ENZYMES

TECHNICAL FIELD

The invention relates to novel nucleic acids, enzymes encoded by these nucleic acids, and to the use of these nucleic acids and proteins in the diagnosis, treatment, and prevention of autoimmune/inflammatory disorders, infectious disorders, immune deficiencies, disorders of metabolism, reproductive disorders, neurological disorders, cardiovascular disorders, eye disorders, and cell proliferative disorders, including cancer. The invention also relates to the assessment of the effects of exogenous compounds on the expression of nucleic acids and enzymes.

BACKGROUND OF THE INVENTION

The cellular processes of biogenesis and biodegradation involve a number of key enzyme classes including oxidoreductases, transferases, hydrolases, lyases, isomerases, ligases, and others. Each class of enzyme comprises many substrate-specific enzymes having precise and well regulated functions. Enzymes facilitate metabolic processes such as glycolysis, the tricarboxylic cycle, and fatty acid metabolism; synthesis or degradation of amino acids, steroids, phospholipids, and alcohols; regulation of cell signaling, proliferation, inflamation, and apoptosis; and through catalyzing critical steps in DNA replication and repair and the process of translation.

Oxidoreductases

Many pathways of biogenesis and biodegradation require oxidoreductase (dehydrogenase or reductase) activity, coupled to reduction or oxidation of a cofactor. Potential cofactors include cytochromes, oxygen, disulfide, iron-sulfur proteins, flavin adenine dinucleotide (FAD), and the nicotinamide adenine dinucleotides NAD and NADP (Newsholme, E.A. and A.R. Leech (1983) Biochemistry for the Medical Sciences, John Wiley and Sons, Chichester, U. K. pp. 779-793). Reductase activity catalyzes transfer of electrons between substrate(s) and cofactor(s) with concurrent oxidation of the cofactor. Reverse dehydrogenase activity catalyzes the reduction of a cofactor and consequent oxidation of the substrate. Oxidoreductase enzymes are a broad superfamily that catalyze reactions in all cells of organisms, including metabolism of sugar, certain detoxification reactions, and synthesis or degradation of fatty acids, amino acids, glucocorticoids, estrogens, androgens, and prostaglandins. Different family members may be referred to as oxidoreductases, oxidases, reductases, or dehydrogenases, and they often have distinct cellular locations such as the cytosol, the plasma membrane, mitochondrial inner or outer membrane, and peroxisomes.

Short-chain alcohol dehydrogenases (SCADs) are a family of dehydrogenases that share only 15% to 30% sequence identity, with similarity predominantly in the coenzyme binding domain and the substrate binding domain. In addition to their role in detoxification of ethanol, SCADs are

involved in synthesis and degradation of fatty acids, steroids, and some prostaglandins, and are therefore implicated in a variety of disorders such as lipid storage disease, myopathy, SCAD deficiency, and certain genetic disorders. For example, retinol dehydrogenase is a SCAD-family member (Simon, A. et al. (1995) J. Biol. Chem. 270:1107-1112) that converts retinol to retinal, the precursor of retinoic acid. Retinoic acid, a regulator of differentiation and apoptosis, has been shown to down-regulate genes involved in cell proliferation and inflammation (Chai, X. et al. (1995) J. Biol. Chem. 270:3900-3904). In addition, retinol dehydrogenase has been linked to hereditary eye diseases such as autosomal recessive childhood-onset severe retinal dystrophy (Simon, A. et al. (1996) Genomics 36:424-430).

Membrane-bound succinate dehydrogenases (succinate:quinone reductases, SQR) and fumarate reductases (quinol:fumarate reductases, QFR) couple the oxidation of succinate to fumarate with the reduction of quinone to quinol, and also catalyze the reverse reaction. QFR and SQR complexes are collectively known as succinate:quinone oxidoreductases (EC 1.3.5.1) and have similar compositions. The complexes consist of two hydrophilic and one or two hydrophobic, membrane-integrated subunits. The larger hydrophilic subunit A carries covalently bound flavin adenine dinucleotide; subunit B contains three iron-sulphur centers (Lancaster, C.R. and A. Kroger (2000) Biochim. Biophys. Acta 1459:422-431). The full-length cDNA sequence for the flavoprotein subunit of human heart succinate dehydrogenase (succinate: (acceptor) oxidoreductase; EC 1.3.99.1) is similar to the bovine succinate dehydrogenase in that it contains a cysteine triplet and in that the active site contains an additional cysteine that is not present in yeast or prokaryotic SQRs (Morris, A.A. et al. (1994) Biochim. Biophys. Acta 29:125-128).

Propagation of nerve impulses, modulation of cell proliferation and differentiation, induction of the immune response, and tissue homeostasis involve neurotransmitter metabolism (Weiss, B. (1991) Neurotoxicology 12:379-386; Collins, S.M. et al. (1992) Ann. N.Y. Acad. Sci. 664:415-424; Brown, J.K. and H. Imam (1991) J. Inherit. Metab. Dis. 14:436-458). Many pathways of neurotransmitter metabolism require oxidoreductase activity, coupled to reduction or oxidation of a cofactor, such as NAD^+/NADH (Newsholme and Leech, *supra*, pp. 779-793). Degradation of catecholamines (epinephrine or norepinephrine) requires alcohol dehydrogenase (in the brain) or aldehyde dehydrogenase (in peripheral tissue). NAD^+-dependent aldehyde dehydrogenase oxidizes 5-hydroxyindole-3-acetate (the product of 5-hydroxytryptamine (serotonin) metabolism) in the brain, blood platelets, liver and pulmonary endothelium (Newsholme and Leech, *supra*, p. 786). Other neurotransmitter degradation pathways that utilize NAD^+/NADH-dependent oxidoreductase activity include those of L-DOPA (precursor of dopamine, a neuronal excitatory compound), glycine (an inhibitory neurotransmitter in the brain and spinal cord), histamine (liberated from mast cells during the inflammatory response), and taurine (an inhibitory neurotransmitter of the brain stem, spinal cord

and retina) (Newsholme and Leech, *supra*, pp. 790, 792). Epigenetic or genetic defects in neurotransmitter metabolic pathways can result in diseases including Parkinson disease and inherited myoclonus (McCance, K.L. and S.E. Huether (1994) Pathophysiology, Mosby-Year Book, Inc., St. Louis, MO pp. 402-404; Gundlach, A.L. (1990) FASEB J. 4:2761-2766).

Tetrahydrofolate is a derivatized glutamate molecule that acts as a carrier, providing activated one-carbon units to a wide variety of biosynthetic reactions, including synthesis of purines, pyrimidines, and the amino acid methionine. Tetrahydrofolate is generated by the activity of a holoenzyme complex called tetrahydrofolate synthase, which includes three enzyme activities: tetrahydrofolate dehydrogenase, tetrahydrofolate cyclohydrolase, and tetrahydrofolate synthetase. Thus, tetrahydrofolate dehydrogenase plays an important role in generating building blocks for nucleic and amino acids, crucial to proliferating cells.

3-Hydroxyacyl-CoA dehydrogenase (3HACD) is involved in fatty acid metabolism. It catalyzes the reduction of 3-hydroxyacyl-CoA to 3-oxoacyl-CoA, with concomitant oxidation of NAD to NADH, in the mitochondria and peroxisomes of eukaryotic cells. In peroxisomes, 3HACD and enoyl-CoA hydratase form an enzyme complex called bifunctional enzyme, defects in which are associated with peroxisomal bifunctional enzyme deficiency. This interruption in fatty acid metabolism produces accumulation of very-long chain fatty acids, disrupting development of the brain, bone, and adrenal glands. Infants born with this deficiency typically die within 6 months (Watkins, P. et al. (1989) J. Clin. Invest. 83:771-777; Online Mendelian Inheritance in Man (OMIM), #261515). The neurodegeneration characteristic of Alzheimer's disease involves development of extracellular plaques in certain brain regions. A major protein component of these plaques is the peptide amyloid-β (Aβ), which is one of several cleavage products of amyloid precursor protein (APP). 3HACD has been shown to bind the Aβ peptide, and is overexpressed in neurons affected in Alzheimer's disease. In addition, an antibody against 3HACD can block the toxic effects of Aβ in a cell culture model of Alzheimer's disease (Yan, S. et al. (1997) Nature 389:689-695; OMIM, #602057).

Steroids such as estrogen, testosterone, and corticosterone are generated from a common precursor, cholesterol, and interconverted. Enzymes acting upon cholesterol include dehydrogenases. Steroid dehydrogenases, such as the hydroxysteroid dehydrogenases, are involved in hypertension, fertility, and cancer (Duax, W.L. and D. Ghosh (1997) Steroids 62:95-100). One such dehydrogenase is 3-oxo-5-α-steroid dehydrogenase (OASD), a microsomal membrane protein highly expressed in prostate and other androgen-responsive tissues. OASD catalyzes the conversion of testosterone into dihydrotestosterone, which is the most potent androgen. Dihydrotestosterone is essential for the formation of the male phenotype during embryogenesis, as well as for proper androgen-mediated growth of tissues such as the prostate and male genitalia. A defect in OASD leads

to defective formation of the external genitalia (Andersson, S. et al. (1991) Nature 354:159-161; Labrie, F. et al. (1992) Endocrinology 131:1571-1573; OMIM #264600).

17β-hydroxysteroid dehydrogenase (17βHSD6) plays an important role in the regulation of the male reproductive hormone, dihydrotestosterone (DHTT). 17βHSD6 acts to reduce levels of DHTT by oxidizing a precursor of DHTT, 3α-diol, to androsterone which is readily glucuronidated and removed. 17βHSD6 is active with both androgen and estrogen substrates in embryonic kidney 293 cells. Isozymes of 17βHSD catalyze oxidation and/or reduction reactions in various tissues with preferences for different steroid substrates (Biswas, M.G. and D.W. Russell (1997) J. Biol. Chem. 272:15959-15966). For example, 17βHSD1 preferentially reduces estradiol and is abundant in the ovary and placenta. 17βHSD2 catalyzes oxidation of androgens and is present in the endometrium and placenta. 17βHSD3 is exclusively a reductive enzyme in the testis (Geissler, W.M. et al. (1994) Nature Genet. 7:34-39). An excess of androgens such as DHTT can contribute to diseases such as benign prostatic hyperplasia and prostate cancer.

The oxidoreductase isocitrate dehydrogenase catalyzes the conversion of isocitrate to a-ketoglutarate, a substrate of the citric acid cycle. Isocitrate dehydrogenase can be either NAD or NADP dependent, and is found in the cytosol, mitochondria, and peroxisomes. Activity of isocitrate dehydrogenase is regulated developmentally, and by hormones, neurotransmitters, and growth factors.

Hydroxypyruvate reductase (HPR), a peroxisomal 2-hydroxyacid dehydrogenase in the glycolate pathway, catalyzes the conversion of hydroxypyruvate to glycerate with the oxidation of both NADH and NADPH. The reverse dehydrogenase reaction reduces NAD^+ and $NADP^+$. HPR recycles nucleotides and bases back into pathways leading to the synthesis of ATP and GTP, which are used to produce DNA and RNA and to control various aspects of signal transduction and energy metabolism. Purine nucleotide biosynthesis inhibitors are used as antiproliferative agents to treat cancer and viral diseases. HPR also regulates biochemical synthesis of serine and cellular serine levels available for protein synthesis.

The mitochondrial electron transport (or respiratory) chain is the series of oxidoreductase-type enzyme complexes in the mitochondrial membrane that is responsible for the transport of electrons from NADH to oxygen and the coupling of this oxidation to the synthesis of ATP (oxidative phosphorylation). ATP provides energy to drive energy-requiring reactions. The key respiratory chain complexes are NADH:ubiquinone oxidoreductase (complex I), succinate:ubiquinone oxidoreductase (complex II), cytochrome c_1-b oxidoreductase (complex III), cytochrome c oxidase (complex IV), and ATP synthase (complex V) (Alberts, B. et al. (1994) Molecular Biology of the Cell, Garland Publishing, Inc., New York, NY, pp. 677-678). All of these complexes are located on the inner matrix side of the mitochondrial membrane except complex II, which is on the cytosolic side where it transports electrons generated in the citric acid cycle to the respiratory chain. Electrons

released in oxidation of succinate to fumarate in the citric acid cycle are transferred through electron carriers in complex II to membrane bound ubiquinone (Q). Transcriptional regulation of these nuclear-encoded genes controls the biogenesis of respiratory enzymes. Defects and altered expression of enzymes in the respiratory chain are associated with a variety of disease conditions.

Other dehydrogenase activities using NAD as a cofactor include 3-hydroxyisobutyrate dehydrogenase (3HBD), which catalyzes the NAD-dependent oxidation of 3-hydroxyisobutyrate to methylmalonate semialdehyde within mitochondria. 3-hydroxyisobutyrate levels are elevated in ketoacidosis, methylmalonic acidemia, and other disorders (Rougraff, P.M. et al. (1989) J. Biol. Chem. 264:5899-5903). Another mitochondrial dehydrogenase important in amino acid metabolism is the enzyme isovaleryl-CoA-dehydrogenase (IVD). IVD is involved in leucine metabolism and catalyzes the oxidation of isovaleryl-CoA to 3-methylcrotonyl-CoA. Human IVD is a tetrameric flavoprotein synthesized in the cytosol with a mitochondrial import signal sequence. A mutation in the gene encoding IVD results in isovaleric acidemia (Vockley, J. et al. (1992) J. Biol. Chem. 267:2494-2501).

The family of glutathione peroxidases encompass tetrameric glutathione peroxidases (GPx1-3) and the monomeric phospholipid hydroperoxide glutathione peroxidase (PHGPx/GPx4). Although the overall homology between the tetrameric enzymes and GPx4 is less than 30%, a pronounced similarity has been detected in clusters involved in the active site and a common catalytic triad has been defined by structural and kinetic data (Epp, O. et al. (1983) Eur. J. Biochem. 133:51-69). GPx1 is ubiquitously expressed in cells, whereas GPx2 is present in the liver and colon, and GPx3 is present in plasma. GPx4 is found at low levels in all tissues but is expressed at high levels in the testis (Ursini, F. et al (1995) Meth. Enzymol. 252:38-53). GPx4 is the only monomeric glutathione peroxidase found in mammals and the only mammalian glutathione peroxidase to show high affinity for and reactivity with phospholipid hydroperoxides, and to be membrane associated. A tandem mechanism for the antioxidant activities of GPx4 and vitamin E has been suggested. GPx4 has alternative transcription and translation start sites which determine its subcellular localization (Esworthy, R.S. et al. (1994) Gene 144:317-318; and Maiorino, M. et al. (1990) Meth. Enzymol. 186:448-450).

The glutathione S-transferases (GST) are a ubiquitous family of enzymes with dual substrate specificities that perform important biochemical functions of xenobiotic biotransformation and detoxification, drug metabolism, and protection of tissues against peroxidative damage. They catalyze the conjugation of an electrophile with reduced glutathione (GSH) which results in either activation or deactivation/detoxification. The absolute requirement for binding reduced GSH to a variety of chemicals necessitates a diversity in GST structures in various organisms and cell types. GSTs are homodimeric or heterodimeric proteins localized in the cytosol. The major isozymes share

common structural and catalytic properties and include four major classes, Alpha, Mu, Pi, and Theta. Each GST possesses a common binding site for GSH, and a variable hydrophobic binding site specific for its particular electrophilic substrates. Specific amino acid residues within GSTs have been identified as important for these binding sites and for catalytic activity. Residues Q67, T68, D101, E104, and R131 are important for the binding of GSH (Lee, H.-C. et al. (1995) J. Biol. Chem. 270:99-109). Residues R13, R20, and R69 are important for the catalytic activity of GST (Stenberg, G. et al. (1991) Biochem. J. 274:549-555).

GSTs normally deactivate and detoxify potentially mutagenic and carcinogenic chemicals. Some forms of rat and human GSTs are reliable preneoplastic markers of carcinogenesis. Dihalomethanes, which produce liver tumors in mice, are believed to be activated by GST (Thier, R. et al. (1993) Proc. Natl. Acad. Sci. USA 90:8567-8580). The mutagenicity of ethylene dibromide and ethylene dichloride is increased in bacterial cells expressing the human Alpha GST, A1-1, while the mutagenicity of aflatoxin B1 is substantially reduced by enhancing the expression of GST (Simula, T.P. et al. (1993) Carcinogenesis 14:1371-1376). Thus, control of GST activity may be useful in the control of mutagenesis and carcinogenesis.

GST has been implicated in the acquired resistance of many cancers to drug treatment, the phenomenon known as multi-drug resistance (MDR). MDR occurs when a cancer patient is treated with a cytotoxic drug such as cyclophosphamide and subsequently becomes resistant to this drug and to a variety of other cytotoxic agents as well. Increased GST levels are associated with some drug resistant cancers, and it is believed that this increase occurs in response to the drug agent which is then deactivated by the GST catalyzed GSH conjugation reaction. The increased GST levels then protect the cancer cells from other cytotoxic agents for which GST has affinity. Increased levels of A1-1 in tumors has been linked to drug resistance induced by cyclophosphamide treatment (Dirven, H.A. et al. (1994) Cancer Res. 54:6215-6220). Thus control of GST activity in cancerous tissues may be useful in treating MDR in cancer patients.

The reduction of ribonucleotides to the corresponding deoxyribonucleotides, needed for DNA synthesis during cell proliferation, is catalyzed by the enzyme ribonucleotide diphosphate reductase. Glutaredoxin is a glutathione (GSH)-dependent hydrogen donor for ribonucleotide diphosphate reductase and contains the active site consensus sequence -C-P-Y-C-. This sequence is conserved in glutaredoxins from such different organisms as *Escherichia coli*, vaccinia virus, yeast, plants, and mammalian cells. Glutaredoxin has inherent GSH-disulfide oxidoreductase (thioltransferase) activity in a coupled system with GSH, NADPH, and GSH-reductase, catalyzing the reduction of low molecular weight disulfides as well as proteins. Glutaredoxin has been proposed to exert a general thiol redox control of protein activity by acting both as an effective protein disulfide reductase, similar to thioredoxin, and as a specific GSH-mixed disulfide reductase (Padilla, C.A. et al. (1996)

FEBS Lett. 378:69-73).

In addition to their important role in DNA synthesis and cell division, glutaredoxin and other thioproteins provide effective antioxidant defense against oxygen radicals and hydrogen peroxide (Schallreuter, K.U. and J.M. Wood (1991) Melanoma Res. 1:159-167). Glutaredoxin is the principal agent responsible for protein dethiolation *in vivo* and reduces dehydroascorbic acid in normal human neutrophils (Jung, C.H. and J.A. Thomas (1996) Arch. Biochem. Biophys. 335:61-72; Park, J.B. and M. Levine (1996) Biochem. J. 315:931-938).

The thioredoxin system serves as a hydrogen donor for ribonucleotide reductase and as a regulator of enzymes by redox control. It also modulates the activity of transcription factors such as NF-κB, AP-1, and steroid receptors. Several cytokines or secreted cytokine-like factors such as adult T-cell leukemia-derived factor, 3B6-interleukin-1, T-hybridoma-derived (MP-6) B cell stimulatory factor, and early pregnancy factor have been reported to be identical to thioredoxin (Holmgren, A. (1985) Annu. Rev. Biochem. 54:237-271; Abate, C. et al. (1990) Science 249:1157-1161; Tagaya, Y. et al. (1989) EMBO J. 8:757-764; Wakasugi, H. (1987) Proc. Natl. Acad. Sci. USA 84:804-808; Rosen, A. et al. (1995) Int. Immunol. 7:625-633). Thus thioredoxin secreted by stimulated lymphocytes (Yodoi, J. and T. Tursz (1991) Adv. Cancer Res. 57:381-411; Tagaya, N. et al. (1990) Proc. Natl. Acad. Sci. USA 87:8282-8286) has extracellular activities including a role as a regulator of cell growth and a mediator in the immune system (Miranda-Vizuete, A. et al. (1996) J. Biol. Chem. 271:19099-19103; Yamauchi, A. et al. (1992) Mol. Immunol. 29:263-270). Thioredoxin and thioredoxin reductase protect against cytotoxicity mediated by reactive oxygen species in disorders such as Alzheimer's disease (Lovell, M.A. (2000) Free Radic. Biol. Med. 28:418-427).

The selenoprotein thioredoxin reductase is secreted by both normal and neoplastic cells and has been implicated as both a growth factor and as a polypeptide involved in apoptosis (Soderberg, A. et al. (2000) Cancer Res. 60:2281-2289). An extracellular plasmin reductase secreted by hamster ovary cells (HT-1080) has been shown to participate in the generation of angiostatin from plasmin. In this case, the reduction of the plasmin disulfide bonds triggers the proteolytic cleavage of plasmin which yields the angiogenesis inhibitor, angiostatin (Stathakis, P. et al. (1997) J. Biol. Chem. 272:20641-20645). Low levels of reduced sulfhydryl groups in plasma has been associated with rheumatoid arthritis. The failure of these sulfhydryl groups to scavenge active oxygen species (e.g., hydrogen peroxide produced by activated neutrophils) results in oxidative damage to surrounding tissues and the resulting inflammation (Hall, N.D. et al. (1994) Rheumatol. Int. 4:35-38).

Another example of the importance of redox reactions in cell metabolism is the degradation of saturated and unsaturated fatty acids by mitochondrial and peroxisomal beta-oxidation enzymes which sequentially remove two-carbon units from Coenzyme A (CoA)-activated fatty acids. The main beta-oxidation pathway degrades both saturated and unsaturated fatty acids while the auxiliary

pathway performs additional steps required for the degradation of unsaturated fatty acids.

The pathways of mitchondrial and peroxisomal beta-oxidation use similar enzymes, but have different substrate specificities and functions. Mitochondria oxidize short-, medium-, and long-chain fatty acids to produce energy for cells. Mitochondrial beta-oxidation is a major energy source for cardiac and skeletal muscle. In liver, it provides ketone bodies to the peripheral circulation when glucose levels are low as in starvation, endurance exercise, and diabetes (Eaton, S. et al. (1996) Biochem. J. 320:345-357). Peroxisomes oxidize medium-, long-, and very-long-chain fatty acids, dicarboxylic fatty acids, branched fatty acids, prostaglandins, xenobiotics, and bile acid intermediates. The chief roles of peroxisomal beta-oxidation are to shorten toxic lipophilic carboxylic acids to facilitate their excretion and to shorten very-long-chain fatty acids prior to mitochondrial beta-oxidation (Mannaerts, G.P. and P.P. Van Veldhoven (1993) Biochimie 75:147-158).

The auxiliary beta-oxidation enzyme 2,4-dienoyl-CoA reductase catalyzes the following reaction:

trans-2, cis/trans-4-dienoyl-CoA + NADPH + H^+ ---> trans-3-enoyl-CoA + $NADP^+$

This reaction removes even-numbered double bonds from unsaturated fatty acids prior to their entry into the main beta-oxidation pathway (Koivuranta, K.T. et al. (1994) Biochem. J. 304:787-792). The enzyme may also remove odd-numbered double bonds from unsaturated fatty acids (Smeland, T.E. et al. (1992) Proc. Natl. Acad. Sci. USA 89:6673-6677).

Rat 2,4-dienoyl-CoA reductase is located in both mitochondria and peroxisomes (Dommes, V. et al. (1981) J. Biol. Chem. 256:8259-8262). Two immunologically different forms of rat mitochondrial enzyme exist with molecular masses of 60 kDa and 120 kDa (Hakkola, E.H. and J.K. Hiltunen (1993) Eur. J. Biochem. 215:199-204). The 120 kDa mitochondrial rat enzyme is synthesized as a 335 amino acid precursor with a 29 amino acid N-terminal leader peptide which is cleaved to form the mature enzyme (Hirose, A. et al. (1990) Biochim. Biophys. Acta 1049:346-349). A human mitochondrial enzyme 83% similar to rat enzyme is synthesized as a 335 amino acid residue precursor with a 19 amino acid N-terminal leader peptide (Koivuranta et al., *supra*). These cloned human and rat mitochondrial enzymes function as homotetramers (Koivuranta et al., *supra*). A *Saccharomyces cerevisiae* peroxisomal 2,4-dienoyl-CoA reductase is 295 amino acids long, contains a C-terminal peroxisomal targeting signal, and functions as a homodimer (Coe, J.G.S. et al. (1994) Mol. Gen. Genet. 244:661-672; and Gurvitz, A. et al. (1997) J. Biol. Chem. 272:22140-22147). All 2,4-dienoyl-CoA reductases have a fairly well conserved NADPH binding site motif (Koivuranta et al., *supra*).

The main pathway beta-oxidation enzyme enoyl-CoA hydratase catalyzes the reaction:

2-trans-enoyl-CoA + H_2O <--> 3-hydroxyacyl-CoA

This reaction hydrates the double bond between C-2 and C-3 of 2-trans-enoyl-CoA, which is generated from saturated and unsaturated fatty acids (Engel, C.K. et al. (1996) EMBO J. 15:5135-5145). This step is downstream from the step catalyzed by 2,4-dienoyl-reductase. Different enoyl-CoA hydratases act on short-, medium-, and long-chain fatty acids (Eaton et al., *supra*). Mitochondrial and peroxisomal enoyl-CoA hydratases occur as both mono-functional enzymes and as part of multi-functional enzyme complexes. Human liver mitochondrial short-chain enoyl-CoA hydratase is synthesized as a 290 amino acid precursor with a 29 amino acid N-terminal leader peptide (Kanazawa, M. et al. (1993) Enzyme Protein 47:9-13; and Janssen, U. et al. (1997) Genomics 40:470-475). Rat short-chain enoyl-CoA hydratase is 87% identical to the human sequence in the mature region of the protein and functions as a homohexamer (Kanazawa et al., *supra*; and Engel et al., *supra*). A mitochondrial trifunctional protein exists that has long-chain enoyl-CoA hydratase, 3-hydroxyacyl-CoA dehydrogenase, and long-chain 3-oxothiolase activities (Eaton et al., *supra*). In human peroxisomes, enoyl-CoA hydratase activity is found in both a 327 amino acid residue mono-functional enzyme and as part of a multi-functional enzyme, also known as bifunctional enzyme, which possesses enoyl-CoA hydratase, enoyl-CoA isomerase, and 3-hydroxyacyl-CoA hydrogenase activities (FitzPatrick, D.R. et al. (1995) Genomics 27:457-466; and Hoefler, G. et al. (1994) Genomics 19:60-67). A 339 amino acid residue human protein with short-chain enoyl-CoA hydratase activity also acts as an AU-specific RNA binding protein (Nakagawa, J. et al. (1995) Proc. Natl. Acad. Sci. USA 92:2051-2055). All enoyl-CoA hydratases share homology near two active site glutamic acid residues, with 17 amino acid residues that are highly conserved (Wu, W.-J. et al. (1997) Biochemistry 36:2211-2220).

Inherited deficiencies in mitochondrial and peroxisomal beta-oxidation enzymes are associated with severe diseases, some of which manifest soon after birth and lead to death within a few years. Mitochondrial beta-oxidation associated deficiencies include, e.g., carnitine palmitoyl transferase and carnitine deficiency, very-long-chain acyl-CoA dehydrogenase deficiency, medium-chain acyl-CoA dehydrogenase deficiency, short-chain acyl-CoA dehydrogenase deficiency, electron transport flavoprotein and electron transport flavoprotein:ubiquinone oxidoreductase deficiency, trifunctional protein deficiency, and short-chain 3-hydroxyacyl-CoA dehydrogenase deficiency (Eaton et al., *supra*). Mitochondrial trifunctional protein (including enoyl-CoA hydratase) deficient patients have reduced long-chain enoyl-CoA hydratase activities and suffer from non-ketotic hypoglycemia, sudden infant death syndrome, cardiomyopathy, hepatic dysfunction, and muscle weakness, and may die at an early age (Eaton et al., *supra*).

Defects in mitochondrial beta-oxidation are associated with Reye's syndrome, a disease

characterized by hepatic dysfunction and encephalopathy that sometimes follows viral infection in children. Reye's syndrome patients may have elevated serum levels of free fatty acids (Cotran, R.S. et al. (1994) Robbins Pathologic Basis of Disease, W.B. Saunders Co., Philadelphia PA, p.866). Patients with mitochondrial short-chain 3-hydroxyacyl-CoA dehydrogenase deficiency and medium-chain 3-hydroxyacyl-CoA dehydrogenase deficiency also exhibit Reye-like illnesses (Eaton et al., *supra*; and Egidio, R.J. et al. (1989) Am. Fam. Physician 39:221-226).

Inherited conditions associated with peroxisomal beta-oxidation include Zellweger syndrome, neonatal adrenoleukodystrophy, infantile Refsum's disease, acyl-CoA oxidase deficiency, peroxisomal thiolase deficiency, and bifunctional protein deficiency (Suzuki, Y. et al. (1994) Am. J. Hum. Genet. 54:36-43; Hoefler et al., *supra*). Patients with peroxisomal bifunctional enzyme deficiency, including that of enoyl-CoA hydratase, suffer from hypotonia, seizures, psychomotor defects, and defective neuronal migration; accumulate very-long-chain fatty acids; and typically die within a few years of birth (Watkins, P.A. et al. (1989) J. Clin. Invest. 83:771-777).

Peroxisomal beta-oxidation is impaired in cancerous tissue. Although neoplastic human breast epithelial cells have the same number of peroxisomes as do normal cells, fatty acyl-CoA oxidase activity is lower than in control tissue (el Bouhtoury, F. et al. (1992) J. Pathol. 166:27-35). Human colon carcinomas have fewer peroxisomes than normal colon tissue and have lower fatty-acyl-CoA oxidase and bifunctional enzyme (including enoyl-CoA hydratase) activities than normal tissue (Cable, S. et al. (1992) Virchows Arch. B Cell Pathol. Incl. Mol. Pathol. 62:221-226).

6-phosphogluconate dehydrogenase (6-PGDH) catalyses the $NADP^+$-dependent oxidative decarboxylation of 6-phosphogluconate to ribulose 5-phosphate with the production of NADPH. The absence or inhibition of 6-PGDH results in the accumulation of 6-phosphogluconate to toxic levels in eukaryotic cells. 6-PGDH is the third enzyme of the pentose phosphate pathway (PPP) and is ubiquitous in nature. In some heterofermentatative species, NAD+ is used as a cofactor with the subsequent production of NADH.

The reaction proceeds through a 3-keto intermediate which is decarboxylated to give the enol of ribulose 5-phosphate, then converted to the keto product following tautomerization of the enol (Berdis A.J. and P.F. Cook (1993) Biochemistry 32:2041-2046). 6-PGDH activity is regulated by the inhibitory effect of NADPH, and the activating effect of 6-phosphogluconate (Rippa, M. et al. (1998) Biochim. Biophys. Acta 1429:83-92). Deficiencies in 6-PGDH activity have been linked to chronic hemolytic anemia.

The targeting of specific forms of 6-PGDH (e.g., enzymes found in trypanosomes) has been suggested as a means for controlling parasitic infections (Tetaud, E. et al. (1999) Biochem. J. 338:55-60). For example, the *Trypanosoma brucei* enzyme is markedly more sensitive to inhibition by the substrate analogue 6-phospho-2-deoxygluconate and the coenzyme analogue adenosine

2',5'-bisphosphate, compared to the mammalian enzyme (Hanau, S. et al. (1996) Eur. J. Biochem. 240:592-599).

Ribonucleotide diphosphate reductase catalyzes the reduction of ribonucleotide diphosphates (i.e., ADP, GDP, CDP, and UDP) to their corresponding deoxyribonucleotide diphosphates (i.e., dADP, dGDP, dCDP, and dUDP) which are used for the synthesis of DNA. Ribonucleotide diphosphate reductase thereby performs a crucial role in the *de novo* synthesis of deoxynucleotide precursors. Deoxynucleotides are also produced from deoxynucleosides by nucleoside kinases via the salvage pathway.

Mammalian ribonucleotide diphosphate reductase comprises two components, an effector-binding component (E) and a non-heme iron component (F). Component E binds the nucleoside triphosphate effectors while component F contains the iron radical necessary for catalysis. Molecular weight determinations of the E and F components, as well as the holoenzyme, vary according to the methods used in purification of the proteins and the particular laboratory. Component E is approximately 90-100 kDa, component F is approximately 100-120 kDa, and the holoenzyme is 200-250 kDa.

Ribonucleotide diphosphate reductase activity is adversely effected by iron chelators, such as thiosemicarbazones, as well as EDTA. Deoxyribonucleotide diphosphates also appear to be negative allosteric effectors of ribonucleotide diphosphate reductase. Nucleotide triphosphates (both ribo- and deoxyribo-) appear to stimulate the activity of the enzyme. 3-methyl-4-nitrophenol, a metabolite of widely used organophosphate pesticides, is a potent inhibitor of ribonucleotide diphosphate reductase in mammalian cells. Some evidence suggests that ribonucleotide diphosphate reductase activity in DNA virus (e.g., herpes virus) -infected cells and in cancer cells is less sensitive to regulation by allosteric regulators and a correlation exists between high ribonucleotide diphosphate reductase activity levels and high rates of cell proliferation (e.g., in hepatomas). This observation suggests that virus-encoded ribonucleotide diphosphate reductases, and those present in cancer cells, are capable of maintaining an increased supply deoxyribonucleotide pool for the production of virus genomes or for the increased DNA synthesis which characterizes cancers cells. Ribonucleotide diphosphate reductase is thus a target for therapeutic intervention (Nutter, L.M. and Y.-C. Cheng (1984) Pharmac. Ther. 26:191-207; and Wright, J.A. (1983) Pharmac. Ther. 22:81-102).

Dihydrodiol dehydrogenases (DD) are monomeric, $NAD(P)^+$-dependent, 34-37 kDa enzymes responsible for the detoxification of *trans*-dihydrodiol and *anti*-diol epoxide metabolites of polycyclic aromatic hydrocarbons (PAH) such as benzo[*a*]yrene, benz[*a*]anthracene, 7-methyl-benz[*a*]anthracene, 7,12-dimethyl-benz[*a*]anthracene, chrysene, and 5-methyl-chrysene. In mammalian cells, an environmental PAH toxin such as benzo[*a*]yrene is initially epoxidated by a microsomal cytochrome P450 to yield 7*R*,8*R*-arene-oxide and subsequently (-)-7*R*,8*R*-dihydrodiol ((-

)-*trans*-7,8-dihydroxy-7,8-dihydrobenzo[*a*]pyrene or (-)-*trans*-B[*a*]P-diol) This latter compound is further transformed to the *anti*-diol epoxide of benzo[*a*]pyrene (i.e., (±)-anti-7β,8α-dihydroxy-9α,10α-epoxy-7,8,9,10-tetrahydrobenzo[*a*]pyrene), by the same enzyme or a different enzyme, depending on the species. This resulting *anti*-diol epoxide of benzo[*a*]yrene, or the corresponding derivative from another PAH compound, is highly mutagenic.

DD efficiently oxidizes the precursor of the *anti*-diol epoxide (i.e., *trans*-dihydrodiol) to transient catechols which auto-oxidize to quinones, also producing hydrogen peroxide and semiquinone radicals. This reaction prevents the formation of the highly carcinogenic *anti*-diol. *Anti*-diols are not themselves substrates for DD yet the addition of DD to a sample comprising an *anti*-diol compound results in a significant decrease in the induced mutation rate observed in the Ames test. In this instance, DD is able to bind to and sequester the *anti*-diol, even though it is not oxidized. Whether through oxidation or sequestration, DD plays an important role in the detoxification of metabolites of xenobiotic polycyclic compounds (Penning, T.M. (1993) Chemico-Biological Interactions 89:1-34).

15-oxoprostaglandin 13-reductase (PGR) and 15-hydroxyprostaglandin dehydrogenase (15-PGDH) are enzymes present in the lung that are responsible for degrading circulating prostaglandins. Oxidative catabolism via passage through the pulmonary system is a common means of reducing the concentration of circulating prostaglandins. 15-PGDH oxidizes the 15-hydroxyl group of a variety of prostaglandins to produce the corresponding 15-oxo compounds. The 15-oxo derivatives usually have reduced biological activity compared to the 15-hydroxyl molecule. PGR further reduces the 13,14 double bond of the 15-oxo compound which typically leads to a further decrease in biological activity. PGR is a monomer with a molecular weight of approximately 36 kDa. The enzyme requires NADH or NADPH as a cofactor with a preference for NADH. The 15-oxo derivatives of prostaglandins PGE_1, PGE_2, and $PGE_{2\alpha}$ are all substrates for PGR; however, the non-derivatized prostaglandins (i.e., PGE_1, PGE_2, and $PGE_{2\alpha}$) are not substrates (Ensor, C.M. et al. (1998) Biochem. J. 330:103-108).

15-PGDH and PGR also catalyze the metabolism of lipoxin A_4 (LXA_4). Lipoxins (LX) are autacoids, lipids produced at the sites of localized inflammation, which down-regulate polymorphonuclear leukocyte (PMN) function and promote resolution of localized trauma. Lipoxin production is stimulated by the administration of aspirin in that cells displaying cyclooxygenase II (COX II) that has been acetylated by aspirin and cells that possess 5-lipoxygenase (5-LO) interact and produce lipoxin. 15-PGDH generates 15-oxo-LXA_4 with PGR further converting the 15-oxo compound to 13,14-dihydro-15-oxo-LXA_4 (Clish, C.B. et al. (2000) J. Biol. Chem. 275:25372-25380). This finding suggests a broad substrate specificity of the prostaglandin dehydrogenases and has implications for these enzymes in drug metabolism and as targets for

therapeutic intervention to regulate inflammation.

The GMC (glucose–methanol–choline) oxidoreductase family of enzymes was defined based on sequence alignments of *Drosophila melanogaster* glucose dehydrogenase, *Escherichia coli* choline dehydrogenase, *Aspergillus niger* glucose oxidase, and *Hansenula polymorpha* methanol oxidase. Despite their different sources and substrate specificities, these four flavoproteins are homologous, being characterized by the presence of several distinctive sequence and structural features. Each molecule contains a canonical ADP-binding, beta-alpha-beta mononucleotide-binding motif close to the amino terminus. This fold comprises a four-stranded parallel beta-sheet sandwiched between a three-stranded antiparallel beta-sheet and alpha-helices. Nucleotides bind in similar positions relative to this chain fold (Cavener, D.R. (1992) J. Mol. Biol. 223:811-814; Wierenga, R.K. et al. (1986) J. Mol. Biol. 187:101-107). Members of the GMC oxidoreductase family also share a consensus sequence near the central region of the polypeptide. Additional members of the GMC oxidoreductase family include cholesterol oxidases from *Brevibacterium sterolicum* and *Streptomyces*; and an alcohol dehydrogenase from *Pseudomonas oleovorans* (Cavener, *supra*; Henikoff, S. and J.G. Henikoff (1994) Genomics 19:97-107; van Beilen, J.B. et al. (1992) Mol. Microbiol. 6:3121-3136).

IMP dehydrogenase and GMP reductase are two oxidoreductases which share many regions of sequence similarity. IMP dehydrogenase (EC 1.1.1.205) catalyes the NAD-dependent reduction of IMP (inosine monophosphate) into XMP (xanthine monophosphate) as part of *de novo* GTP biosynthesis (Collart, F.R. and E. Huberman (1988) J. Biol. Chem. 263:15769-15772). GMP reductase catalyzes the NADPH-dependent reductive deamination of GMP into IMP, helping to maintain the intracellular balance of adenine and guanine nucleotides (Andrews, S.C. and J.R. Guest (1988) Biochem. J. 255:35-43).

Pyridine nucleotide-disulphide oxidoreductases are FAD flavoproteins involved in the transfer of reducing equivalents from FAD to a substrate. These flavoproteins contain a pair of redox-active cysteines contained within a consensus sequence which is characteristic of this protein family (Kurlyan, J. et al. (1991) Nature 352:172-174). Members of this family of oxidoreductases include glutathione reductase (EC 1.6.4.2); thioredoxin reductase of higher eukaryotes (EC 1.6.4.5); trypanothione reductase (EC 1.6.4.8); lipoamide dehydrogenase (EC 1.8.1.4), the E3 component of alpha-ketoacid dehydrogenase complexes; and mercuric reductase (EC 1.16.1.1).

Transferases

Transferases are enzymes that catalyze the transfer of molecular groups. The reaction may involve an oxidation, reduction, or cleavage of covalent bonds, and is often specific to a substrate or to particular sites on a type of substrate. Transferases participate in reactions essential to such functions as synthesis and degradation of cell components, and regulation of cell functions including cell signaling, cell proliferation, inflammation, apoptosis, secretion and excretion. Transferases are

involved in key steps in disease processes involving these functions. Transferases are frequently classified according to the type of group transferred. For example, methyl transferases transfer one-carbon methyl groups, amino transferases transfer nitrogenous amino groups, and similarly denominated enzymes transfer aldehyde or ketone, acyl, glycosyl, alkyl or aryl, isoprenyl, saccharyl, phosphorous-containing, sulfur-containing, or selenium-containing groups, as well as small enzymatic groups such as Coenzyme A.

Acyl transferases include peroxisomal carnitine octanoyl transferase, which is involved in the fatty acid beta-oxidation pathway, and mitochondrial carnitine palmitoyl transferases, involved in fatty acid metabolism and transport. Choline O-acetyl transferase catalyzes the biosynthesis of the neurotransmitter acetylcholine. N-acyltransferase enzymes catalyze the transfer of an amino acid conjugate to an activated carboxylic group. Endogenous compounds and xenobiotics are activated by acyl-CoA synthetases in the cytosol, microsomes, and mitochondria. The acyl-CoA intermediates are then conjugated with an amino acid (typically glycine, glutamine, or taurine, but also ornithine, arginine, histidine, serine, aspartic acid, and several dipeptides) by N-acyltransferases in the cytosol or mitochondria to form a metabolite with an amide bond. One well-characterized enzyme of this class is the bile acid-CoA:amino acid N-acyltransferase (BAT) responsible for generating the bile acid conjugates which serve as detergents in the gastrointestinal tract (Falany, C.N. et al. (1994) J. Biol. Chem. 269:19375-19379; Johnson, M.R. et al. (1991) J. Biol. Chem. 266:10227-10233). BAT is also useful as a predictive indicator for prognosis of hepatocellular carcinoma patients after partial hepatectomy (Furutani, M. et al. (1996) Hepatology 24:1441-1445).

Acetyltransferases

Acetyltransferases have been extensively studied for their role in histone acetylation. Histone acetylation results in the relaxing of the chromatin structure in eukaryotic cells, allowing transcription factors to gain access to promoter elements of the DNA templates in the affected region of the genome (or the genome in general). In contrast, histone deacetylation results in a reduction in transcription by closing the chromatin structure and limiting access of transcription factors. To this end, a common means of stimulating cell transcription is the use of chemical agents that inhibit the deacetylation of histones (e.g., sodium butyrate), resulting in a global (albeit artifactual) increase in gene expression. The modulation of gene expression by acetylation also results from the acetylation of other proteins, including but not limited to, p53, GATA-1, MyoD, ACTR, TFIIE, TFIIF and the high mobility group proteins (HMG). In the case of p53, acetylation results in increased DNA binding, leading to the stimulation of transcription of genes regulated by p53. The prototypic histone acetylase (HAT) is Gcn5 from *Saccharomyces cerevisiae*. Gcn5 is a member of a family of acetylases that includes Tetrahymena p55, human Gcn5, and human p300/CBP. Histone acetylation is reviewed in (Cheung, W.L. et al. (2000) Curr. Opin. Cell Biol. 12:326-333 and Berger, S.L (1999)

Curr. Opin. Cell Biol. 11:336-341). Some acetyltransferase enzymes possess the alpha/beta hydrolase fold (Center of Applied Molecular Engineering Inst. of Chemistry and Biochemistry - University of Salzburg, http://predict.sanger.ac.uk/irbm-course97/Docs/ms/) common to several other major classes of enzymes, including but not limited to, acetylcholinesterases and carboxylesterases (Structural Classification of Proteins, http://scop.mrc-lmb.cam.ac.uk/scop/index.html).

N-acetyltransferases are cytosolic enzymes which utilize the cofactor acetyl-coenzyme A (acetyl-CoA) to transfer the acetyl group to aromatic amines and hydrazine containing compounds. In humans, there are two highly similar N-acetyltransferase enzymes, NAT1 and NAT2; mice appear to have a third form of the enzyme, NAT3. The human forms of N-acetyltransferase have independent regulation (NAT1 is widely-expressed, whereas NAT2 is in liver and gut only) and overlapping substrate preferences. Both enzymes appear to accept most substrates to some extent, but NAT1 does prefer some substrates (para-aminobenzoic acid, para-aminosalicylic acid, sulfamethoxazole, and sulfanilamide), while NAT2 prefers others (isoniazid, hydralazine, procainamide, dapsone, aminoglutethimide, and sulfamethazine). A recently isolated human gene, tubedown-1, is homologous to the yeast NAT-1 N-acetyltransferases and encodes a protein associated with acetyltransferase activity. The expression patterns of tubedown-1 suggest that it may be involved in regulating vascular and hematopoietic development (Gendron, R.L. et al. (2000) Dev. Dyn. 218:300-315).

Amino transferases comprise a family of pyridoxal 5'-phosphate (PLP) -dependent enzymes that catalyze transformations of amino acids. Amino transferases play key roles in protein synthesis and degradation, and they contribute to other processes as well. For example, GABA aminotransferase (GABA-T) catalyzes the degradation of GABA, the major inhibitory amino acid neurotransmitter. The activity of GABA-T is correlated to neuropsychiatric disorders such as alcoholism, epilepsy, and Alzheimer's disease (Sherif, F.M. and S.S. Ahmed (1995) Clin. Biochem. 28:145-154). Other members of the family include pyruvate aminotransferase, branched-chain amino acid aminotransferase, tyrosine aminotransferase, aromatic aminotransferase, alanine:glyoxylate aminotransferase (AGT), and kynurenine aminotransferase (Vacca, R.A. et al. (1997) J. Biol. Chem. 272:21932-21937). Kynurenine aminotransferase catalyzes the irreversible transamination of the L-tryptophan metabolite L-kynurenine to form kynurenic acid. The enzyme may also catalyzes the reversible transamination reaction between L-2-aminoadipate and 2-oxoglutarate to produce 2-oxoadipate and L-glutamate. Kynurenic acid is a putative modulator of glutamatergic neurotransmission, thus a deficiency in kynurenine aminotransferase may be associated with pleiotropic effects (Buchli, R. et al. (1995) J. Biol. Chem. 270:29330-29335).

Glycosyl transferases include the mammalian UDP-glucouronosyl transferases, a family of membrane-bound microsomal enzymes catalyzing the transfer of glucouronic acid to lipophilic

substrates in reactions that play important roles in detoxification and excretion of drugs, carcinogens, and other foreign substances. Another mammalian glycosyl transferase, mammalian UDP-galactose-ceramide galactosyl transferase, catalyzes the transfer of galactose to ceramide in the synthesis of galactocerebrosides in myelin membranes of the nervous system. The UDP-glycosyl transferases share a conserved signature domain of about 50 amino acid residues (PROSITE: PDOC00359, http://expasy.hcuge.ch/sprot/prosite.html).

Methyl transferases are involved in a variety of pharmacologically important processes. Nicotinamide N-methyl transferase catalyzes the N-methylation of nicotinamides and other pyridines, an important step in the cellular handling of drugs and other foreign compounds. Phenylethanolamine N-methyl transferase catalyzes the conversion of noradrenalin to adrenalin. 6-O-methylguanine-DNA methyl transferase reverses DNA methylation, an important step in carcinogenesis. Uroporphyrin-III C-methyl transferase, which catalyzes the transfer of two methyl groups from S-adenosyl-L-methionine to uroporphyrinogen III, is the first specific enzyme in the biosynthesis of cobalamin, a dietary enzyme whose uptake is deficient in pernicious anemia. Protein-arginine methyl transferases catalyze the posttranslational methylation of arginine residues in proteins, resulting in the mono- and dimethylation of arginine on the guanidino group. Substrates include histones, myelin basic protein, and heterogeneous nuclear ribonucleoproteins involved in mRNA processing, splicing, and transport. Protein-arginine methyl transferase interacts with proteins upregulated by mitogens, with proteins involved in chronic lymphocytic leukemia, and with interferon, suggesting an important role for methylation in cytokine receptor signaling (Lin, W.-J. et al. (1996) J. Biol. Chem. 271:15034-15044; Abramovich, C. et al. (1997) EMBO J. 16:260-266; and Scott, H. S. et al. (1998) Genomics 48:330-340).

Phospho transferases catalyze the transfer of high-energy phosphate groups and are important in energy-requiring and -releasing reactions. The metabolic enzyme creatine kinase catalyzes the reversible phosphate transfer between creatine/creatine phosphate and ATP/ADP. Glycocyamine kinase catalyzes phosphate transfer from ATP to guanidoacetate, and arginine kinase catalyzes phosphate transfer from ATP to arginine. A cysteine-containing active site is conserved in this family (PROSITE: PDOC00103).

Prenyl transferases are heterodimers, consisting of an alpha and a beta subunit, that catalyze the transfer of an isoprenyl group. The Ras farnesyltransferase (FTase) enzyme transfers a farnesyl moiety from cytosolic farnesylpyrophosphate to a cysteine residue at the carboxyl terminus of the Ras oncogene protein. This modification is required to anchor Ras to the cell membrane so that it can perform its role in signal transduction. FTase inhibitors block Ras function and demonstrate antitumor activity (Buolamwini, J.K. (1999) Curr. Opin. Chem. Biol. 3:500-509). Ftase, which shares structural similarity with geranylgeranyl transferase, or Rab GG transferase, prenylates Rab proteins,

allowing them to perform their roles in regulating vesicle transport (Seabra, M.C. (1996) J. Biol. Chem. 271:14398-14404).

Saccharyl transferases are glycating enzymes involved in a variety of metabolic processes. Oligosaccharyl transferase-48, for example, is a receptor for advanced glycation endproducts, which accumulate in vascular complications of diabetes, macrovascular disease, renal insufficiency, and Alzheimer's disease (Thornalley, P. J. (1998) Cell Mol. Biol. (Noisy-Le-Grand) 44:1013-1023).

Coenzyme A (CoA) transferase catalyzes the transfer of CoA between two carboxylic acids. Succinyl CoA:3-oxoacid CoA transferase, for example, transfers CoA from succinyl-CoA to a recipient such as acetoacetate. Acetoacetate is essential to the metabolism of ketone bodies, which accumulate in tissues affected by metabolic disorders such as diabetes (PROSITE: PDOC00980).

Transglutaminase transferases (Tgases) are Ca^{2+} dependent enzymes capable of forming isopeptide bonds by catalyzing the transfer of the γ-carboxy group from protein-bound glutamine to the ε-amino group of protein-bound lysine residues or other primary amines. Tgases are the enzymes responsible for the cross-linking of cornified envelope (CE), the highly insoluble protein structure on the surface of corneocytes, into a chemically and mechanically resistant protein polymer. Seven known human Tgases have been identified. Individual transglutaminase gene products are specialized in the cross-linking of specific proteins or tissue structures, such as factor XIIIa which stabilizes the fibrin clot in hemostasis, prostrate transglutaminase which functions in semen coagulation, and tissue transglutaminase which is involved in GTP-binding in receptor signaling. Four (Tgases 1, 2, 3, and X) are expressed in terminally differentiating epithelia such as the epidermis. Tgases are critical for the proper cross-linking of the CE as seen in the pathology of patients suffering from one form of the skin diseases referred to as congenital ichthyosis which has been linked to mutations in the keratinocyte transglutaminase (TG_K) gene (Nemes, Z. et al. (1999) Proc. Natl. Acad. Sci. U.S.A. 96:8402-8407, Aeschlimann, D. et al. (1998) J. Biol. Chem. 273:3452-3460.)

Hydrolases

Hydrolases are a class of enzymes that catalyze the cleavage of various covalent bonds in a substrate by the introduction of a molecule of water. The reaction involves a nucleophilic attack by the water molecule's oxygen atom on a target bond in the substrate. The water molecule is split across the target bond, breaking the bond and generating two product molecules. Hydrolases participate in reactions essential to such functions as synthesis and degradation of cell components, and for regulation of cell functions including cell signaling, cell proliferation, inflamation, apoptosis, secretion and excretion. Hydrolases are involved in key steps in disease processes involving these functions. Hydrolytic enzymes, or hydrolases, may be grouped by substrate specificity into classes including phosphatases, peptidases, lysophospholipases, phosphodiesterases, glycosidases,

glyoxalases, aminohydrolases, carboxylesterases, sulfatases, phosphohydrolases, nucleotidases, lysozymes, and many others.

Phosphatases hydrolytically remove phosphate groups from proteins, an energy-providing step that regulates many cellular processes, including intracellular signaling pathways that in turn control cell growth and differentiation, cell-cell contact, the cell cycle, and oncogenesis.

Peptidases, also called proteases, cleave peptide bonds that form the backbone of peptide or protein chains. Proteolytic processing is essential to cell growth, differentiation, remodeling, and homeostasis as well as inflammation and the immune response. Since typical protein half-lives range from hours to a few days, peptidases are continually cleaving precursor proteins to their active form, removing signal sequences from targeted proteins, and degrading aged or defective proteins. Peptidases function in bacterial, parasitic, and viral invasion and replication within a host. Examples of peptidases include trypsin and chymotrypsin (components of the complement cascade and the blood-clotting cascade) lysosomal cathepsins, calpains, pepsin, renin, and chymosin (Beynon, R.J. and J.S. Bond (1994) Proteolytic Enzymes: A Practical Approach, Oxford University Press, New York, NY, pp. 1-5).

Lysophospholipases (LPLs) regulate intracellular lipids by catalyzing the hydrolysis of ester bonds to remove an acyl group, a key step in lipid degradation. Small LPL isoforms, approximately 15-30 kD, function as hydrolases; larger isoforms function both as hydrolases and transacylases. A particular substrate for LPLs, lysophosphatidylcholine, causes lysis of cell membranes. LPL activity is regulated by signaling molecules important in numerous pathways, including the inflammatory response.

The phosphodiesterases catalyze the hydrolysis of one of the two ester bonds in a phosphodiester compound. Phosphodiesterases are therefore crucial to a variety of cellular processes. Phosphodiesterases include DNA and RNA endo- and exo-nucleases, which are essential to cell growth and replication as well as protein synthesis. Endonuclease V (deoxyinosine 3′-endonuclease) is an example of a type II site-specific deoxyribonuclease, a putative DNA repair enzyme that cleaves DNAs containing hypoxanthine, uracil, or mismatched bases. *Escherichia coli* endonuclease V has been shown to cleave DNA containing deoxyxanthosine at the second phosphodiester bond 3′ to deoxyxanthosine, generating a 3′-hydroxyl and a 5′-phosphoryl group at the nick site (He, B. et al. (2000) Mutat. Res. 459:109-114). It has been suggested that *Escherichia coli* endonuclease V plays a role in the removal of deaminated guanine, i.e., xanthine, from DNA, thus helping to protect the cell against the mutagenic effects of nitrosative deamination (Schouten, K.A. and B. Weiss (1999) Mutat. Res. 435:245-254). In eukaryotes, the process of tRNA splicing requires the removal of small tRNA introns that interrupt the anticodon loop 1 base 3′ to the anticodon. This process requires the stepwise action of an endonuclease, a ligase, and a phosphotransferase (Hong, L. et al. (1998) Science

280:279-284). Ribonuclease P (RNase P) is a ubiquitous RNA processing endonuclease that is required for generating the mature tRNA 5′-end during the tRNA splicing process. This is accomplished through the catalysis of the cleavage of P-3′O bonds to produce 5′-phosphate and 3′-hydroxyl end groups at a specific site on pre-tRNA. Catalysis by RNase P is absolutely dependent on divalent cations such as Mg^{2+} or Mn^{2+} (Kurz, J.C. et al. (2000) Curr. Opin. Chem. Biol. 4:553-558). Substrate recognition mechanisms of RNase P are well conserved among eukaryotes and bacteria (Fabbri, S. et al. (1998) Science 280:284-286). In *Saccharomyces cerevisiae*, POP1 ('processing of precursor RNAs') encodes a protein component of both RNase P and RNase MRP, another RNA processing protein. Mutations in yeast POP1 are lethal (Lygerou, Z. et al. (1994) Genes Dev. 8:1423-1433). Another phosphodiesterase, acid sphingomyelinase, hydrolyzes the membrane phospholipid sphingomyelin to ceramide and phosphorylcholine. Phosphorylcholine functions in synthesis of phosphatidylcholine, which is involved in intracellular signaling pathways. Ceramide is an essential precursor for the generation of gangliosides, membrane lipids found in high concentration in neural tissue. Defective acid sphingomyelinase phosphodiesterase leads to Niemann-Pick disease.

Glycosidases catalyze the cleavage of hemiacetyl bonds of glycosides, which are compounds that contain one or more sugar. Mammalian lactase-phlorizin hydrolase, for example, is an intestinal enzyme that splits lactose. Mammalian beta-galactosidase removes the terminal galactose from gangliosides, glycoproteins, and glycosaminoglycans, and deficiency of this enzyme is associated with a gangliosidosis known as Morquio disease type B (PROSITE PCDOC00910). Vertebrate lysosomal alpha-glucosidase, which hydrolyzes glycogen, maltose, and isomaltose, and vertebrate intestinal sucrase-isomaltase, which hydrolyzes sucrose, maltose, and isomaltose, are widely distributed members of this family with highly conserved sequences at their active sites.

The glyoxylase system is involved in gluconeogenesis, the production of glucose from storage compounds in the body. It consists of glyoxylase I, which catalyzes the formation of S-D-lactoylglutathione from methyglyoxal, a side product of triose-phosphate energy metabolism, and glyoxylase II, which hydrolyzes S-D-lactoylglutathione to D-lactic acid and reduced glutathione. Glyoxylases are involved in hyperglycemia, non-insulin-dependent diabetes mellitus, the detoxification of bacterial toxins, and in the control of cell proliferation and microtubule assembly.

NG,NG-dimethylarginine dimethylaminohydrolase (DDAH) is an enzyme that hydrolyzes the endogenous nitric oxide synthase (NOS) inhibitors, NG-monomethyl-arginine and NG,NG-dimethyl-L-arginine, to L-citrulline. Inhibiting DDAH can cause increased intracellular concentration of NOS inhibitors to levels sufficient to inhibit NOS. Therefore, DDAH inhibition may provide a method of NOS inhibition, and changes in the activity of DDAH could play a role in pathophysiological alterations in nitric oxide generation (MacAllister, R.J. et al. (1996) Br. J. Pharmacol. 119:1533-1540). DDAH was found in neurons displaying cytoskeletal abnormalities and oxidative stress in

Alzheimer's disease. In age-matched control cases, DDAH was not found in neurons. This suggests that oxidative stress- and nitric oxide-mediated events play a role in the pathogenesis of Alzheimer's disease (Smith, M.A. et al. (1998) Free Rad. Biol. Med. 25:898-902).

Acyl-CoA thioesterase is another member of the carboxylesterase family (Alexson, S.E. et al. (1993) Eur. J. Biochem. 214:719-727). Evidence suggests that acyl-CoA thioesterase has a regulatory role in steroidogenic tissues (Finkielstein, C. et al. (1998) Eur. J. Biochem. 256:60-66).

The alpha/beta hydrolase protein fold is common to several hydrolases of diverse phylogenetic origin and catalytic function. Enzymes with the alpha/beta hydrolase fold have a common core structure consisting of eight beta-sheets connected by alpha-helices. The most conserved structural feature of this fold is the loops of the nucleophile-histidine-acid catalytic triad. The histidine in the catalytic triad is completely conserved, while the nucleophile and acid loops accommodate more than one type of amino acid (Ollis, D.L. et al. (1992) Protein Eng. 5:197-211).

Sulfatases are members of a highly conserved gene family that share extensive sequence homology and a high degree of structural similarity. Sulfatases catalyze the cleavage of sulfate esters. To perform this function, sulfatases undergo a unique post-translational modification in the endoplasmic reticulum that involves the oxidation of a conserved cysteine residue. A human disorder called multiple sulfatase deficiency is due to a defect in this post-translational modification step, leading to inactive sulfatases (Recksiek, M. et al. (1998) J. Biol. Chem. 273:6096-6103).

Phosphohydrolases are enzymes that hydrolyze phosphate esters. Some phosphohydrolases contain a mutT domain signature sequence. MutT is a protein involved in the GO system responsible for removing an oxidatively damaged form of guanine from DNA. A region of about 40 amino acid residues, found in the N-terminus of mutT, is also found in other proteins, including some phosphohydrolases (PROSITE PDOC00695).

Serine hydrolases are a large functional class of hydrolytic enzymes that contain a serine residue in their active site. This class of enzymes contains proteinases, esterases, and lipases which hydrolyze a variety of substrates and, therefore, have different biological roles. Proteins in this superfamily can be further grouped into subfamilies based on substrate specificity or amino acid similarities (Puente, X.S. and C. Lopez-Otin (1995) J. Biol. Chem. 270:12926-12932).

Neuropathy target esterase (NTE) is an integral membrane protein present in all neurons and in some non-neural-cell types of vertebrates. NTE is involved in a cell-signaling pathway controlling interactions between neurons and accessory glial cells in the developing nervous system. NTE has serine esterase activity and efficiently catalyses the hydrolysis of phenyl valerate (PV) *in vitro*, but its physiological substrate is unknown. NTE is not related to either the major serine esterase family, which includes acetylcholinesterase, nor to any other known serine hydrolases. NTE contains at least two functional domains: an N-terminal putative regulatory domain and a C-terminal effector domain

which contains the esterase activity and is, in part, conserved in proteins found in bacteria, yeast, nematodes and insects. NTE's effector domain contains three predicted transmembrane segments, and the active-site serine residue lies at the center of one of these segments. The isolated recombinant domain shows PV hydrolase activity only when incorporated into phospholipid liposomes. NTE's esterase activity is largely redundant in adult vertebrates, but organophosphates which react with NTE *in vivo* initiate unknown events which lead to a neuropathy with degeneration of long axons. These neuropathic organophosphates leave a negatively charged group covalently attached to the active-site serine residue, which causes a toxic gain of function in NTE (Glynn, P. (1999) Biochem. J. 344:625-631). Further, the *Drosophila* neurodegeneration gene swiss-cheese encodes a neuronal protein involved in glia-neuron interaction and is homologous to the above human NTE (Moser, M. et al. (2000) Mech. Dev. 90:279-282).

Chitinases are chitin-degrading enzymes present in a variety of organisms and participate in processes including cell wall remodeling, defense and catabolism. Chitinase activity has been found in human serum, leukocytes, granulocytes, and in association with fertilized oocytes in mammals (Escott, G.M. (1995) Infect. Immunol. 63:4770-4773; DeSouza, M.M. (1995) Endrocrinology 136:2485-2496). Glycolytic and proteolytic molecules in humans are associated with tissue damage in lung diseases and with increased tumorigenicity and metastatic potential of cancers (Mulligan, M.S. (1993) Proc. Natl. Acad. Sci. 90:11523-11527; Matrisian, L.M. (1991) Am. J. Med. Sci. 302:157-162; Witty, J.P. (1994) Cancer Res. 54:4805-4812). The discovery of a human enzyme with chitinolytic activity is noteworthy given the lack of endogenous chitin in the human body (Raghavan, N. (1994) Infect. Immun. 62:1901-1908). However, there is a group of mammalian proteins that share homology with chitinases from various non-mammalian organisms, such as bacteria, fungi, plants, and insects. The members of this family differ in their ability to hydrolyze chitin or chitin-like substrates. Some of the mammalian members of the family, such as a bovine whey chitotriosidase and human cartilage proteins which do not demonstrate specific chitinolytic activity, are expressed in association with tissue remodeling events (Rejman, J.J. (1988) Biochem. Biophys. Res. Commun. 150:329-334, Nyirkos, P. (1990) Biochem. J. 268:265-268). Elevated levels of human cartilage proteins have been reported in the synovial fluid and cartilage of patients with rheumatoid arthritis, a disease which produces a severe degradation of the cartilage and a proliferation of the synovial membrane in the affected joints (Hakala, B.E. (1993) J. Biol. Chem. 268:25803-25810).

A small subclass of hydrolases acting on ether bonds includes the thioether hydrolases. *S*-adenosyl-L-homocysteine hydrolase, also known as AdoHcyase or SAHH (PROSITE PDOC00603; EC 3.3.1.1), is a thioether hydrolase first described in rat liver extracts as the activity responsible for the reversible hydrolysis of *S*-adenosyl-L-homocysteine (AdoHcy) to adenosine and homocysteine (Sganga, M.W. et al. (1992) PNAS 89:6328-6332). SAHH is a cytosolic enzyme that has been found

in all cells that have been tested, with the exception of *Escherichia coli* and certain related bacteria (Walker, R.D. et al. (1975) Can. J. Biochem. 53:312-319; Shimizu, S. et al. (1988) FEMS Microbiol. Lett. 51:177-180; Shimizu, S. et al. (1984) Eur. J. Biochem. 141:385-392). SAHH activity is dependent on NAD^+ as a cofactor. Deficiency of SAHH is associated with hypermethioninemia (Online Mendelian Inheritance in Man (OMIM) #180960 Hypermethioninemia), a pathologic condition characterized by neonatal cholestasis, failure to thrive, mental and motor retardation, facial dysmorphism with abnormal hair and teeth, and myocaridopathy (Labrune, P. et al. (1990) J. Pediat. 117:220-226).

Another subclass of hydrolases includes those enzymes which act on carbon-nitrogen (C-N) bonds other than peptide bonds. To this subclass belong those enzymes hydrolyzing amides, amidines, and other C-N bonds. This subclass is further subdivided on the basis of substrate specificity such as linear amides, cyclic amides, linear amidines, cyclic amidines, nitriles and other compounds. A hydrolase belonging to the sub-subclass of enzymes acting on the cyclic amidines is adenosine deaminase (ADA). ADA catalyzes the breakdown of adenosine to inosine. ADA is present in many mammalian tissues, including placenta, muscle, lung, stomach, digestive diverticulum, spleen, erythrocytes, thymus, seminal plasma, thyroid, T-cells, bone marrow stem cells, and liver. A subclass of ADAs, ADAR, act on RNA and are classified as RNA editases. An ADAR from *Drosophila*, dADAR, expressed in the developing nervous system, may act on para voltage-gated Na+ channel transcripts in the central nervous system (Palladino, M.J. et al. (2000) RNA 6:1004-1018). ADA deficiency causes profound lymphopenia with severe combined immunodeficiency (SCID). Cells from patients with ADA deficiency contain low, sometimes undetectable, amounts of ADA catalytic activity and ADA protein. ADA deficiency stems from genetic mutations in the ADA gene (Hershfield, M.S. (1998) Semin. Hematol. 4:291-298). Metabolic consequences of ADA deficiency are associated with defects in alveogenesis, pulmonary inflammation, and airway obstruction (Blackburn, M.R. et al. (2000) J. Exp. Med. 192:159-170).

Pancreatic ribonucleases (RNase) are pyrimidine-specific endonucleases found in high quantity in the pancreas of certain mammalian taxa and of some reptiles (Beintema, J.J. et al (1988) Prog. Biophys. Mol. Biol. 51:165-192). Proteins in the mammalian pancreatic RNase superfamily are noncytosolic endonucleases that degrade RNA through a two-step transphosphorolytic-hydrolytic reaction (Beintema, J.J. et al. (1986) Mol. Biol. Evol. 3:262-275). Specifically, the enzymes are involved in endonucleolytic cleavage of 3'-phosphomononucleotides and 3'-phosphooligonucleotides ending in C-P or U-P with 2',3'-cyclic phosphate intermediates. Ribonucleases can unwind the DNA helix by complexing with single-stranded DNA; the complex arises by an extended multi-site cation-anion interaction between lysine and arginine residues of the enzyme and phosphate groups of the nucleotides. Some of the enzymes belonging to this family appear to play a purely digestive role,

whereas others exhibit potent and unusual biological activities (D'Alessio, G. (1993) Trends Cell Biol. 3:106-109). Proteins belonging to the pancreatic RNase family include: bovine seminal vesicle and brain ribonucleases; kidney non-secretory ribonucleases (Beintema, J.J. et al (1986) FEBS Lett. 194:338-343); liver-type ribonucleases (Rosenberg, H.F. et al. (1989) PNAS U.S.A. 86:4460-4464); angiogenin, which induces vascularisation of normal and malignant tissues; eosinophil cationic protein (Hofsteenge, J. et al. (1989) Biochemistry 28:9806-9813), a cytotoxin and helminthotoxin with ribonuclease activity; and frog liver ribonuclease and frog sialic acid-binding lectin. The sequences of pancreatic RNases contain 4 conserved disulfide bonds and 3 amino acid residues involved in the catalytic activity.

ADP-ribosylation is a reversible post-translational protein modification in which an ADP-ribose moiety is transferred from β-NAD to a target amino acid such as arginine or cysteine. ADP-ribosylarginine hydrolases regenerate arginine by removing ADP-ribose from the protein, completing the ADP-ribosylation cycle (Moss, J. et al. (1997) Adv. Exp. Med. Biol. 419:25-33). ADP-ribosylation is a well-known reaction among bacterial toxins. Cholera toxin, for example, disrupts the adenylyl cyclase system by ADP-ribosylating the α-subunit of the stimulatory G-protein, causing an increase in intracellular cAMP (Moss, J. and M. Vaughan (Eds) (1990) ADP-ribosylating Toxins and G-Proteins: Insights into Signal Transduction, American Society for Microbiology, Washington, D.C.). ADP-ribosylation may also have a regulatory function in eukaryotes, affecting such processes as cytoskeletal assembly (Zhou, H. et al. (1996) Arch. Biochem. Biophys. 334:214-222) and cell proliferation in cytotoxic T-cells (Wang, J. et al. (1996) J. Immunol. 156:2819-2827).

Nucleotidases catalyze the formation of free nucleosides from nucleotides. The cytosolic nucleotidase cN-I (5' nucleotidase-I) cloned from pigeon heart catalyzes the formation of adenosine from AMP generated during ATP hydrolysis (Sala-Newby, G.B. et al. (1999) J. Biol. Chem. 274:17789-17793). Increased adenosine concentration is thought to be a signal of metabolic stress, and adenosine receptors mediate effects including vasodilation, decreased stimulatory neuron firing and ischemic preconditioning in the heart (Schrader, J. (1990) Circulation 81:389-391; Rubino, A. et al. (1992) Eur. J. Pharmacol. 220:95-98; de Jong, J.W. et al. (2000) Pharmacol. Ther. 87:141-149). Deficiency of pyrimidine 5'-nucleotidase can result in hereditary hemolytic anemia (OMIM #266120).

The lysozyme c superfamily consists of conventional lysozymes c, calcium-binding lysozymes c, and α-lactalbumin (Prager, E.M. and P. Jolles (1996) EXS 75:9-31). The proteins in this superfamily have 35-40% sequence homology and share a common three-dimensional fold, but can have different functions. Lysozymes c are ubiquitous in a variety of tissues and secretions and can lyse the cell walls of certain bacteria (McKenzie, H.A. (1996) EXS 75:365-409). Alpha-lactalbumin is a metallo-protein that binds calcium and participates in the synthesis of lactose (Iyer, L.K. and P.K. Qasba (1999) Protein Eng. 12:129-139). Alpha-lactalbumin occurs in mammalian milk and colostrum

(McKenzie, *supra*).

Lysozymes catalyze the hydrolysis of certain mucopolysaccharides of bacterial cell walls, specifically, the beta (1-4) glycosidic linkages between N-acetylmuramic acid and N-acetylglucosamine, and cause bacterial lysis. Lysozymes occur in diverse organisms including viruses, birds, and mammals. In humans, lysozymes are found in spleen, lung, kidney, white blood cells, plasma, saliva, milk, tears, and cartilage (OMIM #153450 Lysozyme; Weaver, L.H. et al. (1985) J. Mol. Biol. 184:739-741). Lysozyme c functions in ruminants as a digestive enzyme, releasing proteins from ingested bacterial cells, and may perform the same function in human newborns (Braun, O.H. et al. (1995) Klin. Pediatr. 207:4-7).

The two known forms of lysozymes, chicken-type and goose-type, were originally isolated from chicken and goose egg white, respectively. Chicken-type and goose-type lysozymes have similar three-dimensional structures, but different amino acid sequences (Nakano, T. and T. Graf (1991) Biochim. Biophys. Acta 1090:273-276). In chickens, both forms of lysozyme are found in neutrophil granulocytes (heterophils), but only chicken-type lysozyme is found in egg white. Generally, chicken-type lysozyme mRNA is found in both adherent monocytes and macrophages and nonadherent promyelocytes and granulocytes as well as in cells of the bone marrow, spleen, bursa, and oviduct. Goose-type lysozyme mRNA is found in non-adherent cells of the bone marrow and lung. Several isozymes have been found in rabbits, including leukocytic, gastrointestinal, and possibly lymphoepithelial forms (OMIM #153450, *supra*; Nakano and Graf, *supra*; and GenBank GI 1310929). A human lysozyme gene encoding a protein similar to chicken-type lysozyme has been cloned (Yoshimura, K. et al. (1988) Biochem. Biophys. Res. Commun. 150:794-801). A consensus motif featuring regularly spaced cysteine residues has been derived from the lysozyme C enzymes of various species (PROSITE PS00128). Lysozyme C shares about 40% amino acid sequence identity with α-lactalbumin.

Lysozymes have several disease associations. Lysozymuria is observed in diabetic nephropathy (Shima, M. et al. (1986) Clin. Chem. 32:1818-1822), endemic nephropathy (Bruckner, I. et al. (1978) Med. Interne. 16:117-125), urinary tract infections (Heidegger, H. (1990) Minerva Ginecol. 42:243-250), and acute monocytic leukemia (Shaw, M.T. (1978) Am. J. Hematol. 4:97-103). Nakano and Graf (*supra*) suggested a role for lysozyme in host defense systems. Older rabbits with an inherited lysozyme deficiency show increased susceptibility to infections, such as subcutaneous abscesses (OMIM #153450, *supra*). Human lysozyme gene mutations cause hereditary systemic amyloidosis, a rare autosomal dominant disease in which amyloid deposits form in the viscera, including the kidney, adrenal glands, spleen, and liver. This disease is usually fatal by the fifth decade. The amyloid deposits contain variant forms of lysozyme. Renal amyloidosis is the most common and potentially the most serious form of organ involvement (Pepys, M.B. et al. (1993)

Nature 362:553-557; OMIM #105200 Familial Visceral Amyloidosis; Cotran, R.S. et al. (1994) Robbins Pathologic Basis of Disease, W.B. Saunders Company, Philadelphia PA, pp. 231-238). Increased levels of lysozyme and lactate have been observed in the cerebrospinal fluid of patients with bacterial meningitis (Ponka, A. et al. (1983) Infection 11:129-131). Acute monocytic leukemia is characterized by massive lysozymuria (Den Tandt, W.R. (1988) Int. J. Biochem. 20:713-719).

Lyases

Lyases are a class of enzymes that catalyze the cleavage of C-C, C-O, C-N, C-S, C-(halide), P-O, or other bonds without hydrolysis or oxidation to form two molecules, at least one of which contains a double bond (Stryer, L. (1995) Biochemistry, W.H. Freeman and Co., New York NY, p.620). Under the International Classification of Enzymes (Webb, E. C. (1992) Enzyme Nomenclature 1992: Recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology on the Nomenclature and Classification of Enzymes, Academic Press, San Diego CA), lyases form a distinct class designated by the numeral 4 in the first digit of the enzyme number (i.e., EC 4.x.x.x).

Further classification of lyases reflects the type of bond cleaved as well as the nature of the cleaved group. The group of C-C lyases includes carboxyl-lyases (decarboxylases), aldehyde-lyases (aldolases), oxo-acid-lyases, and other lyases. The C-O lyase group includes hydro-lyases, lyases acting on polysaccharides, and other lyases. The C-N lyase group includes ammonia-lyases, amidine-lyases, amine-lyases (deaminases), and other lyases. Lyases are critical components of cellular biochemistry, with roles in metabolic energy production, including fatty acid metabolism and the tricarboxylic acid cycle, as well as other diverse enzymatic processes.

One important family of lyases are the carbonic anhydrases (CA), also called carbonate dehydratases, which catalyze the hydration of carbon dioxide in the reaction $H_2O + CO_2 \rightleftharpoons HCO_3^- + H^+$. CA accelerates this reaction by a factor of over 10^6 by virtue of a zinc ion located in a deep cleft about 15Å below the protein's surface and co-ordinated to the imidazole groups of three His residues. Water bound to the zinc ion is rapidly converted to HCO_3^-.

Eight enzymatic and evolutionarily related forms of carbonic anhydrase are currently known to exist in humans: three cytosolic isozymes (CAI, CAII, and CAIII), two membrane-bound forms (CAIV and CAVII), a mitochondrial form (CAV), a secreted salivary form (CAVI) and a yet uncharacterized isozyme (PROSITE PDOC00146 Eukaryotic-type carbonic anhydrases signature). Though the isoenzymes CAI, CAII, and bovine CAIII have similar secondary structures and polypeptide-chain folds, CAI has 6 tryptophans, CAII has 7 and CAIII has 8 (Boren, K. et al. (1996) Protein Sci. 5:2479-2484). CAII is the predominant CA isoenzyme in the brain of mammals.

CAs participate in a variety of physiological processes that involve pH regulation, CO_2 and HCO_3^- transport, ion transport, and water and electrolyte balance. For example, CAII contributes to

H^+ secretion by gastric parietal cells, by renal tubular cells, and by osteoclasts that secrete H^+ to acidify the bone-resorbing compartment. In addition, CAII promotes HCO_3^- secretion by pancreatic duct cells, cilary body epithelium, choroid plexus, salivary gland acinar cells, and distal colonal epithelium, thus playing a role in the production of pancreatic juice, aqueous humor, cerebrospinal fluid, and saliva, and contributing to electrolyte and water balance. CAII also promotes CO_2 exchange in proximal tubules in the kidney, in erythrocytes, and in lung. CAIV has roles in several tissues: it facilitates HCO_3^- reabsorption in the kidney; promotes CO_2 flux in tissues including brain, skeletal muscle, and heart muscle; and promotes CO_2 exchange from the blood to the alveoli in the lung. CAVI probably plays a role in pH regulation in saliva, along with CAII, and may have a protective effect in the esophagus and stomach. Mitochondrial CAV appears to play important roles in gluconeogenesis and ureagenesis, based on the effects of CA inhibitors on these pathways. (Sly, W.S. and P.Y. Hu (1995) Ann. Rev. Biochem. 64:375-401.)

A number of disease states are marked by variations in CA activity. Mutations in CAII which lead to CAII deficiency are the cause of osteopetrosis with renal tubular acidosis (OMIM #259730 Osteopetrosis with Renal Tubular Acidosis). The concentration of CAII in the cerebrospinal fluid (CSF) appears to mark disease activity in patients with brain damage. High CA concentrations have been observed in patients with brain infarction. Patients with transient ischemic attack, multiple sclerosis, or epilepsy usually have CAII concentrations in the normal range, but higher CAII levels have been observed in the CSF of those with central nervous system infection, dementia, or trigeminal neuralgia (Parkkila, A.K. et al. (1997) Eur. J. Clin. Invest. 27:392-397). Colonic adenomas and adenocarcinomas have been observed to fail to stain for CA, whereas non-neoplastic controls showed CAI and CAII in the cytoplasm of the columnar cells lining the upper half of colonic crypts. The neoplasms show staining patterns similar to less mature cells lining the base of normal crypts (Gramlich T.L. et al. (1990) Arch. Pathol. Lab. Med. 114:415-419).

Therapeutic interventions in a number of diseases involve altering CA activity. CA inhibitors such as acetazolamide are used in the treatment of glaucoma (Stewart, W.C. (1999) Curr. Opin. Opthamol. 10:99-108), essential tremor and Parkinson's disease (Uitti, R.J. (1998) Geriatrics 53:46-48, 53-57), intermittent ataxia (Singhvi, J.P. et al. (2000) Neurology India 48:78-80), and altitude related illnesses (Klocke, D.L. et al. (1998) Mayo Clin. Proc. 73:988-992).

CA activity can be particularly useful as an indicator of long-term disease conditions, since the enzyme reacts relatively slowly to physiological changes. CAI and zinc concentrations have been observed to decrease in hyperthyroid Graves' disease (Yoshida, K. (1996) Tohoku J. Exp. Med. 178:345-356) and glycosylated CAI is observed in diabetes mellitus (Kondo, T. et al. (1987) Clin. Chim. Acta 166:227-236). A positive correlation has been observed between CAI and CAII reactivity and endometriosis (Brinton, D.A. et al. (1996) Ann. Clin. Lab. Sci. 26:409-420; D'Cruz , O.J. et al.

(1996) Fertil. Steril. 66:547-556).

Another important member of the lyase family is ornithine decarboxylase (ODC), the initial rate-limiting enzyme in polyamine biosynthesis. ODC catalyses the transformation of ornithine into putrescine in the reaction L-ornithine ⇌ putrescine + CO_2. Polyamines, which include putrescine and the subsequent metabolic pathway products spermidine and spermine, are ubiquitous cell components essential for DNA synthesis, cell differentiation, and proliferation. Thus the polyamines play a key role in tumor proliferation (Medina, M.A. et al. (1999) Biochem. Pharmacol. 57:1341-1344).

ODC is a pyridoxal-5'-phosphate (PLP)-dependent enzyme which is active as a homodimer. Conserved residues include those at the PLP binding site and a stretch of glycine residues thought to be part of a substrate binding region (PROSITE PDOC00685 Orn/DAP/Arg decarboxylase family 2 signatures). Mammalian ODCs also contain PEST regions, sequence fragments enriched in proline, glutamic acid, serine, and threonine residues that act as signals for intracellular degradation (Medina et al., *supra*).

Many chemical carcinogens and tumor promoters increase ODC levels and activity. Several known oncogenes may increase ODC levels by enhancing transcription of the ODC gene, and ODC itself may act as an oncogene when expressed at very high levels. A high level of ODC is found in a number of precancerous conditions, and elevation of ODC levels has been used as part of a screen for tumor-promoting compounds (Pegg, A.E. et al. (1995) J. Cell. Biochem. Suppl. 22:132-138).

Inhibitors of ODC have been used to treat tumors in animal models and human clinical trials, and have been shown to reduce development of tumors of the bladder, brain, esophagus, gastrointestinal tract, lung, oral cavity, mammary gland, stomach, skin and trachea (Pegg et al., *supra*; McCann, P.P. and A.E. Pegg (1992) Pharmac. Ther. 54:195-215). ODC also shows promise as a target for chemoprevention (Pegg et al., *supra*). ODC inhibitors have also been used to treat infections by African trypanosomes, malaria, and *Pneumocystis carinii*, and are potentially useful for treatment of autoimmune diseases such as lupus and rheumatoid arthritis (McCann and Pegg, *supra*).

Another family of pyridoxal-dependent decarboxylases are the group II decarboxylases. This family includes glutamate decarboxylase (GAD) which catalyzes the decarboxylation of glutamate into the neurotransmitter GABA; histidine decarboxylase (HDC), which catalyzes the decarboxylation of histidine to histamine; aromatic-L-amino-acid decarboxylase (DDC), also known as L-dopa decarboxylase or tryptophan decarboxylase, which catalyzes the decarboxylation of tryptophan to tryptamine and also acts on 5-hydroxy-tryptophan and dihydroxyphenylalanine (L-dopa); and cysteine sulfinic acid decarboxylase (CSD), the rate-limiting enzyme in the synthesis of taurine from cysteine (PROSITE PDOC00329 DDC/GAD/HDC/TyrDC pyridoxal-phosphate attachment site). Taurine is an abundant sulfonic amino acid in brain and is thought to act as an osmoregulator in brain cells (Bitoun, M. and M. Tappaz (2000) J. Neurochem. 75:919-924).

Isomerases

Isomerases are a class of enzymes that catalyze geometric or structural changes within a molecule to form a single product. This class includes racemases and epimerases, cis-trans-isomerases, intramolecular oxidoreductases, intramolecular transferases (mutases) and intramolecular lyases. Isomerases are critical components of cellular biochemistry with roles in metabolic energy production including glycolysis, as well as other diverse enzymatic processes (Stryer, *supra*, pp.483-507).

Racemases are a subset of isomerases that catalyze inversion of a molecule's configuration around the asymmetric carbon atom in a substrate having a single center of asymmetry, thereby interconverting two racemers. Epimerases are another subset of isomerases that catalyze inversion of configuration around an asymmetric carbon atom in a substrate with more than one center of symmetry, thereby interconverting two epimers. Racemases and epimerases can act on amino acids and derivatives, hydroxy acids and derivatives, and carbohydrates and derivatives. The interconversion of UDP-galactose and UDP-glucose is catalyzed by UDP-galactose-4'-epimerase. Proper regulation and function of this epimerase is essential to the synthesis of glycoproteins and glycolipids. Elevated blood galactose levels have been correlated with UDP-galactose-4'-epimerase deficiency in screening programs of infants (Gitzelmann, R. (1972) Helv. Paediat. Acta 27:125-130).

Correct folding of newly synthesized proteins is assisted by molecular chaperones and folding catalysts, two unrelated groups of helper molecules. Chaperones suppress non-productive side reactions by stoichiometric binding to folding intermediates, whereas folding enzymes catalyze some of the multiple folding steps that enable proteins to attain their final functional configurations (Kern, G. et al. (1994) FEBS Lett. 348:145-148). One class of folding enzymes, the peptidyl prolyl *cis-trans* isomerases (PPIases), isomerizes certain proline imidic bonds in what is considered to be a rate limiting step in protein maturation and export. PPIases catalyze the *cis* to *trans* isomerization of certain proline imidic bonds in proteins. There are three evolutionarily unrelated families of PPIases: the cyclophilins, the FK506 binding proteins, and the newly characterized parvulin family (Rahfeld, J.U. et al. (1994) FEBS Lett. 352:180-184).

The cyclophilins (CyP) were originally identified as major receptors for the immunosuppressive drug cyclosporin A (CsA), an inhibitor of T-cell activation (Handschumacher, R.E. et al. (1984) Science 226:544-547; Harding, M.W. et al. (1986) J. Biol. Chem. 261:8547-8555). Thus, the peptidyl-prolyl isomerase activity of CyP may be part of the signaling pathway that leads to T-cell activation. Subsequent work demonstrated that CyP's isomerase activity is essential for correct protein folding and/or protein trafficking, and may also be involved in assembly/disassembly of protein complexes and regulation of protein activity. For example, in *Drosophila*, the CyP NinaA is required for correct localization of rhodopsins, while a mammalian CyP (Cyp40) is part of the

Hsp90/Hsp70 complex that binds steroid receptors. The mammalian CyP (CypA) has been shown to bind the *gag* protein from human immunodeficiency virus 1 (HIV-1), an interaction that can be inhibited by cyclosporin. Since cyclosporin has potent anti-HIV-1 activity, CypA may play an essential function in HIV-1 replication. Finally, Cyp40 has been shown to bind and inactivate the transcription factor c-Myb, an effect that is reversed by cyclosporin. This effect implicates CyP in the regulation of transcription, transformation, and differentiation (Bergsma, D.J. et al (1991) J. Biol. Chem. 266:23204-23214; Hunter, T. (1998) Cell 92:141-143; and Leverson, J.D. and S.A. Ness (1998) Mol. Cell. 1:203-211).

One of the major rate limiting steps in protein folding is the thiol:disulfide exchange that is necessary for correct protein assembly. Although incubation of reduced, unfolded proteins in buffers with defined ratios of oxidized and reduced thiols can lead to native conformation, the rate of folding is slow and the attainment of native conformation decreases proportionately with the size and number of cysteines in the protein. Certain cellular compartments such as the endoplasmic reticulum of eukaryotes and the periplasmic space of prokaryotes are maintained in a more oxidized state than the surrounding cytosol. Correct disulfide formation can occur in these compartments, but at a rate that is insufficient for normal cell processes and inadequate for synthesizing secreted proteins. The protein disulfide isomerases, thioredoxins and glutaredoxins are able to catalyze the formation of disulfide bonds and regulate the redox environment in cells to enable the necessary thiol:disulfide exchanges (Loferer, H. (1995) J. Biol. Chem. 270:26178-26183).

Each of these proteins has somewhat different functions, but all belong to a group of disulfide-containing redox proteins that contain a conserved active-site sequence and are ubiquitously distributed in eukaryotes and prokaryotes. Protein disulfide isomerases are found in the endoplasmic reticulum of eukaryotes and in the periplasmic space of prokaryotes. They function by exchanging their own disulfide for a thiol in a folding peptide chain. In contrast, the reduced thioredoxins and glutaredoxins are generally found in the cytoplasm and function by directly reducing disulfides in the substrate proteins.

Oxidoreductases can be isomerases as well. Oxidoreductases catalyze the reversible transfer of electrons from a substrate that becomes oxidized to a substrate that becomes reduced. This class of enzymes includes dehydrogenases, hydroxylases, oxidases, oxygenases, peroxidases, and reductases. Proper maintenance of oxidoreductase levels is physiologically important. For example, genetically-linked deficiencies in lipoamide dehydrogenase can result in lactic acidosis (Robinson, B. H. et al. (1977) Pediat. Res. 11:1198-1202).

Another subgroup of isomerases are the transferases (or mutases). Transferases transfer a chemical group from one compound (the donor) to another compound (the acceptor). The types of groups transferred by these enzymes include acyl groups, amino groups, phosphate groups

(phosphotransferases or phosphomutases), and others. The transferase carnitine palmitoyltransferase is an important component of fatty acid metabolism. Genetically-linked deficiencies in this transferase can lead to myopathy (Scriver, C. et al. (1995) The Metabolic and Molecular Basis of Inherited Disease, McGraw-Hill, New York NY, pp.1501-1533).

Yet another subgroup of isomerases are the topoisomersases. Topoisomerases are enzymes that affect the topological state of DNA. For example, defects in topoisomerases or their regulation can affect normal physiology. Reduced levels of topoisomerase II have been correlated with some of the DNA processing defects associated with the disorder ataxia-telangiectasia (Singh, S.P. et al. (1988) Nucleic Acids Res. 16:3919-3929).

Ligases

Ligases catalyze the formation of a bond between two substrate molecules. The process involves the hydrolysis of a pyrophosphate bond in ATP or a similar energy donor. Ligases are classified based on the nature of the type of bond they form, which can include carbon-oxygen, carbon-sulfur, carbon-nitrogen, carbon-carbon and phosphoric ester bonds.

Ligases forming carbon-oxygen bonds include the aminoacyl-transfer RNA (tRNA) synthetases which are important RNA-associated enzymes with roles in translation. Protein biosynthesis depends on each amino acid forming a linkage with the appropriate tRNA. The aminoacyl-tRNA synthetases are responsible for the activation and correct attachment of an amino acid with its cognate tRNA. The 20 aminoacyl-tRNA synthetase enzymes can be divided into two structural classes, and each class is characterized by a distinctive topology of the catalytic domain. Class I enzymes contain a catalytic domain based on the nucleotide-binding "Rossman fold". Class II enzymes contain a central catalytic domain, which consists of a seven-stranded antiparallel ß-sheet motif, as well as N- and C- terminal regulatory domains. Class II enzymes are separated into two groups based on the heterodimeric or homodimeric structure of the enzyme; the latter group is further subdivided by the structure of the N- and C-terminal regulatory domains (Hartlein, M. and S. Cusack, (1995) J. Mol. Evol. 40:519-530). Autoantibodies against aminoacyl-tRNAs are generated by patients with dermatomyositis and polymyositis, and correlate strongly with complicating interstitial lung disease (ILD). These antibodies appear to be generated in response to viral infection, and coxsackie virus has been used to induce experimental viral myositis in animals.

Ligases forming carbon-sulfur bonds (acid-thiol ligases) mediate a large number of cellular biosynthetic intermediary metabolism processes involving intermolecular transfer of carbon atom-containing substrates (carbon substrates). Examples of such reactions include the tricarboxylic acid cycle, synthesis of fatty acids and long-chain phospholipids, synthesis of alcohols and aldehydes, synthesis of intermediary metabolites, and reactions involved in the amino acid degradation pathways. Some of these reactions require input of energy, usually in the form of conversion of ATP

to either ADP or AMP and pyrophosphate.

In many cases, a carbon substrate is derived from a small molecule containing at least two carbon atoms. The carbon substrate is often covalently bound to a larger molecule which acts as a carbon substrate carrier molecule within the cell. In the biosynthetic mechanisms described above, the carrier molecule is coenzyme A. Coenzyme A (CoA) is structurally related to derivatives of the nucleotide ADP and consists of 4'-phosphopantetheine linked via a phosphodiester bond to the alpha phosphate group of adenosine 3',5'-bisphosphate. The terminal thiol group of 4'-phosphopantetheine acts as the site for carbon substrate bond formation. The predominant carbon substrates which utilize CoA as a carrier molecule during biosynthesis and intermediary metabolism in the cell are acetyl, succinyl, and propionyl moieties, collectively referred to as acyl groups. Other carbon substrates include enoyl lipid, which acts as a fatty acid oxidation intermediate, and carnitine, which acts as an acetyl-CoA flux regulator/mitochondrial acyl group transfer protein. Acyl-CoA and acetyl-CoA are synthesized in the cell by acyl-CoA synthetase and acetyl-CoA synthetase, respectively.

Activation of fatty acids is mediated by at least three forms of acyl-CoA synthetase activity: i) acetyl-CoA synthetase, which activates acetate and several other low molecular weight carboxylic acids and is found in muscle mitochondria and the cytosol of other tissues; ii) medium-chain acyl-CoA synthetase, which activates fatty acids containing between four and eleven carbon atoms (predominantly from dietary sources), and is present only in liver mitochondria; and iii) acyl CoA synthetase, which is specific for long chain fatty acids with between six and twenty carbon atoms, and is found in microsomes and the mitochondria. Proteins associated with acyl-CoA synthetase activity have been identified from many sources including bacteria, yeast, plants, mouse, and man. The activity of acyl-CoA synthetase may be modulated by phosphorylation of the enzyme by cAMP-dependent protein kinase.

Ligases forming carbon-nitrogen bonds include amide synthases such as glutamine synthetase (glutamate-ammonia ligase) that catalyzes the amination of glutamic acid to glutamine by ammonia using the energy of ATP hydrolysis. Glutamine is the primary source for the amino group in various amide transfer reactions involved in *de novo* pyrimidine nucleotide synthesis and in purine and pyrimidine ribonucleotide interconversions. Overexpression of glutamine synthetase has been observed in primary liver cancer (Christa, L. et al. (1994) Gastroent. 106:1312-1320).

Acid-amino-acid ligases (peptide synthases) are represented by the ubiquitin conjugating enzymes which are associated with the ubiquitin conjugation system (UCS), a major pathway for the degradation of cellular proteins in eukaryotic cells and some bacteria. The UCS mediates the elimination of abnormal proteins and regulates the half-lives of important regulatory proteins that control cellular processes such as gene transcription and cell cycle progression. In the UCS pathway, proteins targeted for degradation are conjugated to ubiquitin (Ub), a small heat stable protein. Ub is

first activated by a ubiquitin-activating enzyme (E1), and then transferred to one of several Ub-conjugating enzymes (E2). E2 then links the Ub molecule through its C-terminal glycine to an internal lysine (acceptor lysine) of a target protein. The ubiquitinated protein is then recognized and degraded by proteasome, a large, multisubunit proteolytic enzyme complex, and ubiquitin is released for reutilization by ubiquitin protease. The UCS is implicated in the degradation of mitotic cyclic kinases, oncoproteins, tumor suppressor genes such as p53, viral proteins, cell surface receptors associated with signal transduction, transcriptional regulators, and mutated or damaged proteins (Ciechanover, A. (1994) Cell 79:13-21).

Cyclo-ligases and other carbon-nitrogen ligases comprise various enzymes and enzyme complexes that participate in the *de novo* pathways of purine and pyrimidine biosynthesis. Because these pathways are critical to the synthesis of nucleotides for replication of both RNA and DNA, many of these enzymes have been the targets of clinical agents for the treatment of cell proliferative disorders such as cancer and infectious diseases.

Purine biosynthesis occurs *de novo* from the amino acids glycine and glutamine, and other small molecules. Three of the key reactions in this process are catalyzed by a trifunctional enzyme composed of glycinamide-ribonucleotide synthetase (GARS), aminoimidazole ribonucleotide synthetase (AIRS), and glycinamide ribonucleotide transformylase (GART). Together these three enzymes combine ribosylamine phosphate with glycine to yield phosphoribosyl aminoimidazole, a precursor to both adenylate and guanylate nucleotides. This trifunctional protein has been implicated in the pathology of Downs syndrome (Aimi, J. et al. (1990) Nucleic Acid Res. 18:6665-6672). Adenylosuccinate synthetase catalyzes a later step in purine biosynthesis that converts inosinic acid to adenylosuccinate, a key step on the path to ATP synthesis. This enzyme is also similar to another carbon-nitrogen ligase, argininosuccinate synthetase, that catalyzes a similar reaction in the urea cycle (Powell, S.M. et al. (1992) FEBS Lett. 303:4-10).

Adenylosuccinate synthetase, adenylosuccinate lyase, and AMP deaminase may be considered as a functional unit, the purine nucleotide cycle. This cycle converts AMP to inosine monophosphate (IMP) and reconverts IMP to AMP via adenylosuccinate, thereby producing NH_3 and forming fumarate from aspartate. In muscle, the purine nucleotide cycle functions, during intense exercise, in the regeneration of ATP by pulling the adenylate kinase reaction in the direction of ATP formation and by providing Krebs cycle intermediates. In kidney, the purine nucleotide cycle accounts for the release of NH_3 under normal acid-base conditions. In brain, the purine nucleotide cycle may contribute to ATP recovery. Adenylosuccinate lyase deficiency provokes psychomotor retardation, often accompanied by autistic features (Van den Berghe, G. et al. (1992) Prog Neurobiol.: 39:547-561). A marked imbalance in the enzymic pattern of purine metabolism is linked with transformation and/or progression in cancer cells. In rat hepatomas the specific activities of the

anabolic enzymes, IMP dehydrogenase, GMP synthetase, adenylosuccinate synthetase, adenylosuccinase, AMP deaminase and amidophosphoribosyltransferase, increased to 13.5-, 3.7-, 3.1-, 1.8-, 5.5- and 2.8-fold, respectively, of those in normal liver (Weber, G. (1983) Clin. Biochem. 16:57-63).

Like the *de novo* biosynthesis of purines, *de novo* synthesis of the pyrimidine nucleotides uridylate and cytidylate also arises from a common precursor, in this instance the nucleotide orotidylate derived from orotate and phosphoribosyl pyrophosphate (PPRP). Again a trifunctional enzyme comprising three carbon-nitrogen ligases plays a key role in the process. In this case the enzymes aspartate transcarbamylase (ATCase), carbamyl phosphate synthetase II, and dihydroorotase (DHOase) are encoded by a single gene called CAD. Together these three enzymes combine the initial reactants in pyrimidine biosynthesis, glutamine, CO_2, and ATP to form dihydroorotate, the precursor to orotate and orotidylate (Iwahana, H. et al. (1996) Biochem. Biophys. Res. Commun. 219:249-255). Further steps then lead to the synthesis of uridine nucleotides from orotidylate. Cytidine nucleotides are derived from uridine-5'-triphosphate (UTP) by the amidation of UTP using glutamine as the amino donor and the enzyme CTP synthetase. Regulatory mutations in the human CTP synthetase are believed to confer multi-drug resistance to agents widely used in cancer therapy (Yamauchi, M. et al. (1990) EMBO J. 9:2095-2099).

Ligases forming carbon-carbon bonds include the carboxylases acetyl-CoA carboxylase and pyruvate carboxylase. Acetyl-CoA carboxylase catalyzes the carboxylation of acetyl-CoA from CO_2 and H_2O using the energy of ATP hydrolysis. Acetyl-CoA carboxylase is the rate-limiting enzyme in the biogenesis of long-chain fatty acids. Two isoforms of acetyl-CoA carboxylase, types I and types II, are expressed in human in a tissue-specific manner (Ha, J. et al. (1994) Eur. J. Biochem. 219:297-306). Pyruvate carboxylase is a nuclear-encoded mitochondrial enzyme that catalyzes the conversion of pyruvate to oxaloacetate, a key intermediate in the citric acid cycle.

Ligases forming phosphoric ester bonds include the DNA ligases involved in both DNA replication and repair. DNA ligases seal phosphodiester bonds between two adjacent nucleotides in a DNA chain using the energy from ATP hydrolysis to first activate the free 5'-phosphate of one nucleotide and then react it with the 3'-OH group of the adjacent nucleotide. This resealing reaction is used in DNA replication to join small DNA fragments called "Okazaki" fragments that are transiently formed in the process of replicating new DNA, and in DNA repair. DNA repair is the process by which accidental base changes, such as those produced by oxidative damage, hydrolytic attack, or uncontrolled methylation of DNA, are corrected before replication or transcription of the DNA can occur. Bloom's syndrome is an inherited human disease in which individuals are partially deficient in DNA ligation and consequently have an increased incidence of cancer (Alberts et al., *supra*, p. 247).

Pantothenate synthetase (D-pantoate; beta-alanine ligase (AMP-forming); EC 6.3.2.1) is the last enzyme of the pathway of pantothenate (vitamin B(5)) synthesis. It catalyzes the condensation of pantoate with beta-alanine in an ATP-dependent reaction. The enzyme is dimeric, with two well-defined domains per protomer: the N-terminal domain, a Rossmann fold, contains the active site cavity, with the C-terminal domain forming a hinged lid. The N-terminal domain is structurally very similar to class I aminoacyl-tRNA synthetases and is thus a member of the cytidylyltransferase superfamily (von Delft, F. et al. (2000) Structure (Camb) 9:439-450).

Farnesyl diphosphate synthase (FPPS) is an essential enzyme that is required both for cholesterol synthesis and protein prenylation. The enzyme catalyzes the formation of farnesyl diphosphate from dimethylallyl diphosphate and isopentyl diphosphate. FPPS is inhibited by nitrogen-containing biphosphonates, which can lead to the inhibition of osteoclast-mediated bone resorption by preventing protein prenylation (Dunford, J.E. et al. (2001) J. Pharmacol. Exp. Ther. 296:235-242).

5-aminolevulinate synthase (ALAS; delta-aminolevulinate synthase; EC 2.3.1.37) catalyzes the rate-limiting step in heme biosynthesis in both erythroid and non-erythroid tissues. This enzyme is unique in the heme biosynthetic pathway in being encoded by two genes, the first encoding ALAS1, the non-erythroid specific enzyme which is ubiquitously expressed, and the second encoding ALAS2, which is expressed exclusively in erythroid cells. The genes for ALAS1 and ALAS2 are located, respectively, on chromosome 3 and on the X chromosome. Defects in the gene encoding ALAS2 result in X-linked sideroblastic anemia. Elevated levels of ALAS are seen in acute hepatic porphyrias and can be lowered by zinc mesoporphyrin.

Drug Metabolizing Enzymes (DMEs)

The metabolism of a drug and its movement through the body (pharmacokinetics) are important in determining its effects, toxicity, and interactions with other drugs. The three processes governing pharmacokinetics are the absorption of the drug, distribution to various tissues, and elimination of drug metabolites. These processes are intimately coupled to drug metabolism, since a variety of metabolic modifications alter most of the physicochemical and pharmacological properties of drugs, including solubility, binding to receptors, and excretion rates. The metabolic pathways which modify drugs also accept a variety of naturally occurring substrates such as steroids, fatty acids, prostaglandins, leukotrienes, and vitamins. The enzymes in these pathways are therefore important sites of biochemical and pharmacological interaction between natural compounds, drugs, carcinogens, mutagens, and xenobiotics. It has long been appreciated that inherited differences in drug metabolism lead to drastically different levels of drug efficacy and toxicity among individuals. Advances in pharmacogenomics research, of which DMEs constitute an important part, are promising to expand the tools and information that can be brought to bear on questions of drug efficacy and

toxicity (See Evans, W.E. and R.V. Relling (1999) Science 286:487-491). DMEs have broad substrate specificities, unlike antibodies, for example, which are diverse and highly specific. Since DMEs metabolize a wide variety of molecules, drug interactions may occur at the level of metabolism so that, for example, one compound may induce a DME that affects the metabolism of another compound.

Drug metabolic reactions are categorized as Phase I, which prepare the drug molecule for functioning and further metabolism, and Phase II, which are conjugative. In general, Phase I reaction products are partially or fully inactive, and Phase II reaction products are the chief excreted species. However, Phase I reaction products are sometimes more active than the original administered drugs; this metabolic activation principle is exploited by pro-drugs (e.g. L-dopa). Additionally, some nontoxic compounds (e.g. aflatoxin, benzo[*a*]pyrene) are metabolized to toxic intermediates through these pathways. Phase I reactions are usually rate-limiting in drug metabolism. Prior exposure to the compound, or other compounds, can induce the expression of Phase I enzymes however, and thereby increase substrate flux through the metabolic pathways. (See Klaassen, C.D. et al. (1996) Casarett and Doull's Toxicology: The Basic Science of Poisons, McGraw-Hill, New York, NY, pp. 113-186; Katzung, B.G. (1995) Basic and Clinical Pharmacology, Appleton and Lange, Norwalk, CT, pp. 48-59; Gibson, G.G. and P. Skett (1994) Introduction to Drug Metabolism, Blackie Academic and Professional, London.).

The major classes of Phase I enzymes include, but are not limited to, cytochrome P450 and flavin-containing monooxygenase. Other enzyme classes involved in Phase I-type catalytic cycles and reactions include, but are not limited to, NADPH cytochrome P450 reductase (CPR), the microsomal cytochrome b5/NADH cytochrome b5 reductase system, the ferredoxin/ferredoxin reductase redox pair, aldo/keto reductases, and alcohol dehydrogenases. The major classes of Phase II enzymes include, but are not limited to, UDP glucuronyltransferase, sulfotransferase, glutathione S-transferase, N-acyltransferase, and N-acetyl transferase.

Cytochrome P450 and P450 catalytic cycle-associated enzymes

Members of the cytochrome P450 superfamily of enzymes catalyze the oxidative metabolism of a variety of substrates, including natural compounds such as steroids, fatty acids, prostaglandins, leukotrienes, and vitamins, as well as drugs, carcinogens, mutagens, and xenobiotics. Cytochromes P450, also known as P450 heme-thiolate proteins, usually act as terminal oxidases in multi-component electron transfer chains, called P450-containing monooxygenase systems. Specific reactions catalyzed include hydroxylation, epoxidation, N-oxidation, sulfooxidation, N-, S-, and O-dealkylations, desulfation, deamination, and reduction of azo, nitro, and N-oxide groups. These reactions are involved in steroidogenesis of glucocorticoids, cortisols, estrogens, and androgens in animals; insecticide resistance in insects; herbicide resistance and flower coloring in plants; and

environmental bioremediation by microorganisms. Cytochrome P450 actions on drugs, carcinogens, mutagens, and xenobiotics can result in detoxification or in conversion of the substance to a more toxic product. Cytochromes P450 are abundant in the liver, but also occur in other tissues; the enzymes are located in microsomes. (See ExPASY ENZYME EC 1.14.14.1; Prosite PDOC00081 Cytochrome P450 cysteine heme-iron ligand signature; PRINTS EP450I E-Class P450 Group I signature; Graham-Lorence, S. and J.A. Peterson (1996) FASEB J. 10:206-214.)

Four hundred cytochromes P450 have been identified in diverse organisms including bacteria, fungi, plants, and animals (Graham-Lorence and Peterson, *supra*). The B-class is found in prokaryotes and fungi, while the E-class is found in bacteria, plants, insects, vertebrates, and mammals. Five subclasses or groups are found within the larger family of E-class cytochromes P450 (PRINTS EP450I E-Class P450 Group I signature).

All cytochromes P450 use a heme cofactor and share structural attributes. Most cytochromes P450 are 400 to 530 amino acids in length. The secondary structure of the enzyme is about 70% alpha-helical and about 22% beta-sheet. The region around the heme-binding site in the C-terminal part of the protein is conserved among cytochromes P450. A ten amino acid signature sequence in this heme-iron ligand region has been identified which includes a conserved cysteine involved in binding the heme iron in the fifth coordination site. In eukaryotic cytochromes P450, a membrane-spanning region is usually found in the first 15-20 amino acids of the protein, generally consisting of approximately 15 hydrophobic residues followed by a positively charged residue. (See Prosite PDOC00081, *supra*; Graham-Lorence and Peterson, *supra*.)

Cytochrome P450 enzymes are involved in cell proliferation and development. The enzymes have roles in chemical mutagenesis and carcinogenesis by metabolizing chemicals to reactive intermediates that form adducts with DNA (Nebert, D.W. and F.J. Gonzalez (1987) Ann. Rev. Biochem. 56:945-993). These adducts can cause nucleotide changes and DNA rearrangements that lead to oncogenesis. Cytochrome P450 expression in liver and other tissues is induced by xenobiotics such as polycyclic aromatic hydrocarbons, peroxisomal proliferators, phenobarbital, and the glucocorticoid dexamethasone (Dogra, S.C. et al. (1998) Clin. Exp. Pharmacol. Physiol. 25:1-9). A cytochrome P450 protein may participate in eye development as mutations in the P450 gene CYP1B1 cause primary congenital glaucoma (OMIM #601771 Cytochrome P450, subfamily I (dioxin-inducible), polypeptide 1; CYP1B1).

Cytochromes P450 are associated with inflammation and infection. Hepatic cytochrome P450 activities are profoundly affected by various infections and inflammatory stimuli, some of which are suppressed and some induced (Morgan, E.T. (1997) Drug Metab. Rev. 29:1129-1188). Effects observed *in vivo* can be mimicked by proinflammatory cytokines and interferons. Autoantibodies to two cytochrome P450 proteins were found in patients with autoimmune

polyenodocrinopathy-candidiasis-ectodermal dystrophy (APECED), a polyglandular autoimmune syndrome (OMIM #240300 Autoimmune polyenodocrinopathy-candidiasis-ectodermal dystrophy).

Mutations in cytochromes P450 have been linked to metabolic disorders, including congenital adrenal hyperplasia, the most common adrenal disorder of infancy and childhood; pseudovitamin D-deficiency rickets; cerebrotendinous xanthomatosis, a lipid storage disease characterized by progressive neurologic dysfunction, premature atherosclerosis, and cataracts; and an inherited resistance to the anticoagulant drugs coumarin and warfarin (Isselbacher, K.J. et al. (1994) Harrison's Principles of Internal Medicine, McGraw-Hill, Inc. New York, NY, pp. 1968-1970; Takeyama, K. et al. (1997) Science 277:1827-1830; Kitanaka, S. et al. (1998) N. Engl. J. Med. 338:653-661; OMIM #213700 Cerebrotendinous xanthomatosis; and OMIM #122700 Coumarin resistance). Extremely high levels of expression of the cytochrome P450 protein aromatase were found in a fibrolamellar hepatocellular carcinoma from a boy with severe gynecomastia (feminization) (Agarwal, V.R. (1998) J. Clin. Endocrinol. Metab. 83:1797-1800).

The cytochrome P450 catalytic cycle is completed through reduction of cytochrome P450 by NADPH cytochrome P450 reductase (CPR). Another microsomal electron transport system consisting of cytochrome b5 and NADPH cytochrome b5 reductase has been widely viewed as a minor contributor of electrons to the cytochrome P450 catalytic cycle. However, a recent report by Lamb, D.C. et al. (1999; FEBS Lett. 462:283-288) identifies a *Candida albicans* cytochrome P450 (CYP51) which can be efficiently reduced and supported by the microsomal cytochrome b5/NADPH cytochrome b5 reductase system. Therefore, there are likely many cytochromes P450 which are supported by this alternative electron donor system.

Cytochrome b5 reductase is also responsible for the reduction of oxidized hemoglobin (methemoglobin, or ferrihemoglobin, which is unable to carry oxygen) to the active hemoglobin (ferrohemoglobin) in red blood cells. Methemoglobinemia results when there is a high level of oxidant drugs or an abnormal hemoglobin (hemoglobin M) which is not efficiently reduced. Methemoglobinemia can also result from a hereditary deficiency in red cell cytochrome b5 reductase (Reviewed in Mansour, A. and A.A. Lurie (1993) Am. J. Hematol. 42:7-12).

Members of the cytochrome P450 family are also closely associated with vitamin D synthesis and catabolism. Vitamin D exists as two biologically equivalent prohormones, ergocalciferol (vitamin D_2), produced in plant tissues, and cholecalciferol (vitamin D_3), produced in animal tissues. The latter form, cholecalciferol, is formed upon the exposure of 7-dehydrocholesterol to near ultraviolet light (i.e., 290-310 nm), normally resulting from even minimal periods of skin exposure to sunlight (reviewed in Miller, W.L. and A.A. Portale (2000) Trends Endocrinol. Metab. 11:315-319).

Both prohormone forms are further metabolized in the liver to 25-hydroxyvitamin D (25(OH)D) by the enzyme 25-hydroxylase. 25(OH)D is the most abundant precursor form of vitamin

D which must be further metabolized in the kidney to the active form, 1α,25-dihydroxyvitamin D ($1\alpha,25(OH)_2D$), by the enzyme 25-hydroxyvitamin D 1α-hydroxylase (1α-hydroxylase). Regulation of $1\alpha,25(OH)_2D$ production is primarily at this final step in the synthetic pathway. The activity of 1α-hydroxylase depends upon several physiological factors including the circulating level of the enzyme product ($1\alpha,25(OH)_2D$) and the levels of parathyroid hormone (PTH), calcitonin, insulin, calcium, phosphorus, growth hormone, and prolactin. Furthermore, extrarenal 1α-hydroxylase activity has been reported, suggesting that tissue-specific, local regulation of 1α,25(OH)2D production may also be biologically important. The catalysis of $1\alpha,25(OH)_2D$ to 24,25-dihydroxyvitamin D ($24,25(OH)_2D$), involving the enzyme 25-hydroxyvitamin D 24-hydroxylase (24-hydroxylase), also occurs in the kidney. 24-hydroxylase can also use 25(OH)D as a substrate (Shinki, T. et al. (1997) Proc. Natl. Acad. Sci. U.S.A. 94:12920-12925; Miller and Portale, *supra*; and references within).

Vitamin D 25-hydroxylase, 1α-hydroxylase, and 24-hydroxylase are all NADPH-dependent, type I (mitochondrial) cytochrome P450 enzymes that show a high degree of homology with other members of the family. Vitamin D 25-hydroxylase also shows a broad substrate specificity and may also perform 26-hydroxylation of bile acid intermediates and 25, 26, and 27-hydroxylation of cholesterol (Dilworth, F.J. et al. (1995) J. Biol. Chem. 270:16766-16774; Miller and Portale, *supra*; and references within).

The active form of vitamin D ($1\alpha,25(OH)_2D$) is involved in calcium and phosphate homeostasis and promotes the differentiation of myeloid and skin cells. Vitamin D deficiency resulting from deficiencies in the enzymes involved in vitamin D metabolism (e.g., 1α-hydroxylase) causes hypocalcemia, hypophosphatemia, and vitamin D-dependent (sensitive) rickets, a disease characterized by loss of bone density and distinctive clinical features, including bandy or bow leggedness accompanied by a waddling gait. Deficiencies in vitamin D 25-hydroxylase cause cerebrotendinous xanthomatosis, a lipid-storage disease characterized by the deposition of cholesterol and cholestanol in the Achilles' tendons, brain, lungs, and many other tissues. The disease presents with progressive neurologic dysfunction, including postpubescent cerebellar ataxia, atherosclerosis, and cataracts. Vitamin D 25-hydroxylase deficiency does not result in rickets, suggesting the existence of alternative pathways for the synthesis of 25(OH)D (Griffin, J.E. and J.E. Zerwekh (1983) J. Clin. Invest. 72:1190-1199; Gamblin, G.T. et al. (1985) J. Clin. Invest. 75:954-960; and Miller and Portale, *supra*).

Ferredoxin and ferredoxin reductase are electron transport accessory proteins which support at least one human cytochrome P450 species, cytochrome P450c27 encoded by the CYP27 gene (Dilworth, F.J. et al. (1996) Biochem. J. 320:267-71). A *Streptomyces griseus* cytochrome P450, CYP104D1, was heterologously expressed in *Escherichia coli* and found to be reduced by the

endogenous ferredoxin and ferredoxin reductase enzymes (Taylor, M. et al. (1999) Biochem. Biophys. Res. Commun. 263:838-842), suggesting that many cytochrome P450 species may be supported by the ferredoxin/ferredoxin reductase pair. Ferredoxin reductase has also been found in a model drug metabolism system to reduce actinomycin D, an antitumor antibiotic, to a reactive free radical species (Flitter, W.D. and R.P. Mason (1988) Arch. Biochem. Biophys. 267:632-639).

Flavin-containing monooxygenase (FMO)

Flavin-containing monooxygenases oxidize the nucleophilic nitrogen, sulfur, and phosphorus heteroatom of an exceptional range of substrates. Like cytochromes P450, FMOs are microsomal and use NADPH and O_2; there is also a great deal of substrate overlap with cytochromes P450. The tissue distribution of FMOs includes liver, kidney, and lung.

Isoforms of FMO in mammals include FMO1, FMO2, FMO3, FMO4, and FMO5, which are expressed in a tissue-specific manner. The isoforms differ in their substrate specificities and properties such as inhibition by various compounds and stereospecificity of reaction. FMOs have a 13 amino acid signature sequence, the components of which span the N-terminal two-thirds of the sequences and include the FAD binding region and the FATGY motif found in many N-hydroxylating enzymes (Stehr, M. et al. (1998) Trends Biochem. Sci. 23:56-57; PRINTS FMOXYGENASE Flavin-containing monooxygenase signature). Specific reactions include oxidation of nucleophilic tertiary amines to N-oxides, secondary amines to hydroxylamines and nitrones, primary amines to hydroxylamines and oximes, and sulfur-containing compounds and phosphines to S- and P-oxides. Hydrazines, iodides, selenides, and boron-containing compounds are also substrates. FMOs are more heat labile and less detergent-sensitive than cytochromes P450 *in vitro* though FMO isoforms vary in thermal stability and detergent sensitivity.

FMOs play important roles in the metabolism of several drugs and xenobiotics. FMO (FMO3 in liver) is predominantly responsible for metabolizing (S)-nicotine to (S)-nicotine N-1'-oxide, which is excreted in urine. FMO is also involved in S-oxygenation of cimetidine, an H_2-antagonist widely used for the treatment of gastric ulcers. Liver-expressed forms of FMO are not under the same regulatory control as cytochrome P450. In rats, for example, phenobarbital treatment leads to the induction of cytochrome P450, but the repression of FMO1.

Lysyl oxidase

Lysyl oxidase (lysine 6-oxidase, LO) is a copper-dependent amine oxidase involved in the formation of connective tissue matrices by crosslinking collagen and elastin. LO is secreted as an N-glycosylated precursor protein of approximately 50 kDa and cleaved to the mature form of the enzyme by a metalloprotease, although the precursor form is also active. The copper atom in LO is involved in the transport of electrons to and from oxygen to facilitate the oxidative deamination of lysine residues in these extracellular matrix proteins. While the coordination of copper is essential to

LO activity, insufficient dietary intake of copper does not influence the expression of the apoenzyme. However, the absence of the functional LO is linked to the skeletal and vascular tissue disorders that are associated with dietary copper deficiency. LO is also inhibited by a variety of semicarbazides, hydrazines, and amino nitrites, as well as heparin. Beta-aminopropionitrile is a commonly used inhibitor. LO activity is increased in response to ozone, cadmium, and elevated levels of hormones released in response to local tissue trauma, such as transforming growth factor-beta, platelet-derived growth factor, angiotensin II, and fibroblast growth factor. Abnormalities in LO activity have been linked to Menkes syndrome and occipital horn syndrome. Cytosolic forms of the enzyme have been implicated in abnormal cell proliferation (reviewed in Rucker, R.B. et al. (1998) Am. J. Clin. Nutr. 67:996S-1002S and Smith-Mungo, L.I. and H.M. Kagan (1998) Matrix Biol. 16:387-398).

Dihydrofolate reductases

Dihydrofolate reductases (DHFR) are ubiquitous enzymes that catalyze the NADPH-dependent reduction of dihydrofolate to tetrahydrofolate, an essential step in the *de novo* synthesis of glycine and purines as well as the conversion of deoxyuridine monophosphate (dUMP) to deoxythymidine monophosphate (dTMP). The basic reaction is as follows:

$$\text{7,8-dihydrofolate} + \text{NADPH} \rightarrow \text{5,6,7,8-tetrahydrofolate} + \text{NADP}^+$$

The enzymes can be inhibited by a number of dihydrofolate analogs, including trimethroprim and methotrexate. Since an abundance of dTMP is required for DNA synthesis, rapidly dividing cells require the activity of DHFR. The replication of DNA viruses (i.e., herpesvirus) also requires high levels of DHFR activity. As a result, drugs that target DHFR have been used for cancer chemotherapy and to inhibit DNA virus replication. (For similar reasons, thymidylate synthetases are also target enzymes.) Drugs that inhibit DHFR are preferentially cytotoxic for rapidly dividing cells (or DNA virus-infected cells) but have no specificity, resulting in the indiscriminate destruction of dividing cells. Furthermore, cancer cells may become resistant to drugs such as methotrexate as a result of acquired transport defects or the duplication of one or more DHFR genes (Stryer, L. (1988) Biochemistry. W.H Freeman and Co., Inc. New York. pp. 511-519).

Aldo/keto reductases

Aldo/keto reductases are monomeric NADPH-dependent oxidoreductases with broad substrate specificities (Bohren, K.M. et al. (1989) J. Biol. Chem. 264:9547-9551). These enzymes catalyze the reduction of carbonyl-containing compounds, including carbonyl-containing sugars and aromatic compounds, to the corresponding alcohols. Therefore, a variety of carbonyl-containing drugs and xenobiotics are likely metabolized by enzymes of this class.

One known reaction catalyzed by a family member, aldose reductase, is the reduction of

glucose to sorbitol, which is then further metabolized to fructose by sorbitol dehydrogenase. Under normal conditions, the reduction of glucose to sorbitol is a minor pathway. In hyperglycemic states, however, the accumulation of sorbitol is implicated in the development of diabetic complications (OMIM #103880 Aldo-keto reductase family 1, member B1). Members of this enzyme family are also highly expressed in some liver cancers (Cao, D. et al. (1998) J. Biol. Chem. 273:11429-11435).

Alcohol dehydrogenases

Alcohol dehydrogenases (ADHs) oxidize simple alcohols to the corresponding aldehydes. ADH is a cytosolic enzyme, prefers the cofactor NAD^+, and also binds zinc ion. Liver contains the highest levels of ADH, with lower levels in kidney, lung, and the gastric mucosa.

Known ADH isoforms are dimeric proteins composed of 40 kDa subunits. There are five known gene loci which encode these subunits (a, b, g, p, c), and some of the loci have characterized allelic variants (b_1, b_2, b_3, g_1, g_2). The subunits can form homodimers and heterodimers; the subunit composition determines the specific properties of the active enzyme. The holoenzymes have therefore been categorized as Class I (subunit compositions aa, ab, ag, bg, gg), Class II (pp), and Class III (cc). Class I ADH isozymes oxidize ethanol and other small aliphatic alcohols, and are inhibited by pyrazole. Class II isozymes prefer longer chain aliphatic and aromatic alcohols, are unable to oxidize methanol, and are not inhibited by pyrazole. Class III isozymes prefer even longer chain aliphatic alcohols (five carbons and longer) and aromatic alcohols, and are not inhibited by pyrazole.

The short-chain alcohol dehydrogenases include a number of related enzymes with a variety of substrate specificities. Included in this group are the mammalian enzymes D-beta-hydroxybutyrate dehydrogenase, (R)-3-hydroxybutyrate dehydrogenase, 15-hydroxyprostaglandin dehydrogenase, NADPH-dependent carbonyl reductase, corticosteroid 11-beta-dehydrogenase, and estradiol 17-beta-dehydrogenase, as well as the bacterial enzymes acetoacetyl-CoA reductase, glucose 1-dehydrogenase, 3-beta-hydroxysteroid dehydrogenase, 20-beta-hydroxysteroid dehydrogenase, ribitol dehydrogenase, 3-oxoacyl reductase, 2,3-dihydro-2,3-dihydroxybenzoate dehydrogenase, sorbitol-6-phosphate 2-dehydrogenase, 7-alpha-hydroxysteroid dehydrogenase, cis-1,2-dihydroxy-3,4-cyclohexadiene-1-carboxylate dehydrogenase, *cis*-toluene dihydrodiol dehydrogenase, *cis*-benzene glycol dehydrogenase, biphenyl-2,3-dihydro-2,3-diol dehydrogenase, N-acylmannosamine 1-dehydrogenase, and 2-deoxy-D-gluconate 3-dehydrogenase (Krozowski, Z. (1994) J. Steroid Biochem. Mol. Biol. 51:125-130; Krozowski, Z. (1992) Mol. Cell Endocrinol. 84:C25-31; and Marks, A.R. et al. (1992) J. Biol. Chem. 267:15459-15463).

Sulfotransferases

Sulfate conjugation occurs on many of the same substrates which undergo O-glucuronidation to produce a highly water-soluble sulfuric acid ester. Sulfotransferases (ST) catalyze this reaction by

transferring SO_3^- from the cofactor 3'-phosphoadenosine-5'-phosphosulfate (PAPS) to the substrate. ST substrates are predominantly phenols and aliphatic alcohols, but also include aromatic amines and aliphatic amines, which are conjugated to produce the corresponding sulfamates. The products of these reactions are excreted mainly in urine.

STs are found in a wide range of tissues, including liver, kidney, intestinal tract, lung, platelets, and brain. The enzymes are generally cytosolic, and multiple forms are often co-expressed. For example, there are more than a dozen forms of ST in rat liver cytosol. These biochemically characterized STs fall into five classes based on their substrate preference: arylsulfotransferase, alcohol sulfotransferase, estrogen sulfotransferase, tyrosine ester sulfotransferase, and bile salt sulfotransferase.

ST enzyme activity varies greatly with sex and age in rats. The combined effects of developmental cues and sex-related hormones are thought to lead to these differences in ST expression profiles, as well as the profiles of other DMEs such as cytochromes P450. Notably, the high expression of STs in cats partially compensates for their low level of UDP glucuronyltransferase activity.

Several forms of ST have been purified from human liver cytosol and cloned. There are two phenol sulfotransferases with different thermal stabilities and substrate preferences. The thermostable enzyme catalyzes the sulfation of phenols such as para-nitrophenol, minoxidil, and acetaminophen; the thermolabile enzyme prefers monoamine substrates such as dopamine, epinephrine, and levadopa. Other cloned STs include an estrogen sulfotransferase and an N-acetylglucosamine-6-O-sulfotransferase. This last enzyme is illustrative of the other major role of STs in cellular biochemistry, the modification of carbohydrate structures that may be important in cellular differentiation and maturation of proteoglycans. Indeed, an inherited defect in a sulfotransferase has been implicated in macular corneal dystrophy, a disorder characterized by a failure to synthesize mature keratan sulfate proteoglycans (Nakazawa, K. et al. (1984) J. Biol. Chem. 259:13751-13757; OMIM #217800 Macular dystrophy, corneal).

Galactosyltransferases

Galactosyltransferases are a subset of glycosyltransferases that transfer galactose (Gal) to the terminal N-acetylglucosamine (GlcNAc) oligosaccharide chains that are part of glycoproteins or glycolipids that are free in solution (Kolbinger, F. et al. (1998) J. Biol. Chem. 273:433-440; Amado, M. et al. (1999) Biochim. Biophys. Acta 1473:35-53). Galactosyltransferases have been detected on the cell surface and as soluble extracellular proteins, in addition to being present in the Golgi. β1,3-galactosyltransferases form Type I carbohydrate chains with Gal (β1-3)GlcNAc linkages. Known human and mouse β1,3-galactosyltransferases appear to have a short cytosolic domain, a single transmembrane domain, and a catalytic domain with eight conserved regions. (Kolbinger et al., *supra;*

and Hennet, T. et al. (1998) J. Biol. Chem. 273:58-65). In mouse UDP-galactose:β-N-acetylglucosamine β1,3-galactosyltransferase-I region 1 is located at amino acid residues 78-83, region 2 is located at amino acid residues 93-102, region 3 is located at amino acid residues 116-119, region 4 is located at amino acid residues 147-158, region 5 is located at amino acid residues 172-183, region 6 is located at amino acid residues 203-206, region 7 is located at amino acid residues 236-246, and region 8 is located at amino acid residues 264-275. A variant of a sequence found within mouse UDP-galactose:β-N-acetylglucosamine β1,3-galactosyltransferase-I region 8 is also found in bacterial galactosyltransferases, suggesting that this sequence defines a galactosyltransferase sequence motif (Hennet et al., *supra*). Recent work suggests that brainiac protein is a β1,3-galactosyltransferase (Yuan, Y. et al. (1997) Cell 88:9-11; and Hennet et al., *supra*).

UDP-Gal:GlcNAc-1,4-galactosyltransferase (-1,4-GalT) (Sato, T. et al., (1997) EMBO J. 16:1850-1857) catalyzes the formation of Type II carbohydrate chains with Gal (β1-4)GlcNAc linkages. As is the case with the β1,3-galactosyltransferase, a soluble form of the enzyme is formed by cleavage of the membrane-bound form. Amino acids conserved among β1,4-galactosyltransferases include two cysteines linked through a disulfide-bond and a putative UDP-galactose-binding site in the catalytic domain (Yadav, S. and K. Brew (1990) J. Biol. Chem. 265:14163-14169; Yadav, S.P. and K. Brew (1991) J. Biol. Chem. 266:698-703; and Shaper, N.L. et al. (1997) J. Biol. Chem. 272:31389-31399). β1,4-galactosyltransferases have several specialized roles in addition to synthesizing carbohydrate chains on glycoproteins or glycolipids. In mammals a β1,4-galactosyltransferase, as part of a heterodimer with α-lactalbumin, functions in lactating mammary gland lactose production. A β1,4-galactosyltransferase on the surface of sperm functions as a receptor that specifically recognizes the egg. Cell surface β1,4-galactosyltransferases also function in cell adhesion, cell/basal lamina interaction, and normal and metastatic cell migration. (Shur, B. (1993) Curr. Opin. Cell Biol. 5:854-863; and Shaper, J. (1995) Adv. Exp. Med. Biol. 376:95-104).

Gamma-glutamyl transpeptidase

Gamma-glutamyl transpeptidases are ubiquitously expressed enzymes that initiate extracellular glutathione (GSH) breakdown by cleaving gamma-glutamyl amide bonds. The breakdown of GSH provides cells with a regional cysteine pool for biosynthetic pathways. Gamma-glutamyl transpeptidases also contribute to cellular antioxidant defenses and expression is induced by oxidative stress. The cell surface-localized glycoproteins.are expressed at high levels in cancer cells. Studies have suggested that the high level of gamma-glutamyl transpeptidase activity present on the surface of cancer cells could be exploited to activate precursor drugs, resulting in high local concentrations of anti-cancer therapeutic agents (Hanigan, M.H. (1998) Chem. Biol. Interact. 111-112:333-342; Taniguchi, N. and Y. Ikeda (1998) Adv. Enzymol. Relat. Areas Mol. Biol. 72:239-278; Chikhi, N. et al. (1999) Comp. Biochem. Physiol. B. Biochem. Mol. Biol. 122:367-380).

Aminotransferases

Aminotransferases comprise a family of pyridoxal 5'-phosphate (PLP) -dependent enzymes that catalyze transformations of amino acids. Aspartate aminotransferase (AspAT) is the most extensively studied PLP-containing enzyme. It catalyzes the reversible transamination of dicarboxylic L-amino acids, aspartate and glutamate, and the corresponding 2-oxo acids, oxalacetate and 2-oxoglutarate. Other members of the family include pyruvate aminotransferase, branched-chain amino acid aminotransferase, tyrosine aminotransferase, aromatic aminotransferase, alanine:glyoxylate aminotransferase (AGT), and kynurenine aminotransferase (Vacca, R.A. et al. (1997) J. Biol. Chem. 272:21932-21937).

Primary hyperoxaluria type-1 is an autosomal recessive disorder resulting in a deficiency in the liver-specific peroxisomal enzyme, alanine:glyoxylate aminotransferase-1. The phenotype of the disorder is a deficiency in glyoxylate metabolism. In the absence of AGT, glyoxylate is oxidized to oxalate rather than being transaminated to glycine. The result is the deposition of insoluble calcium oxalate in the kidneys and urinary tract, ultimately causing renal failure (Lumb, M.J. et al. (1999) J. Biol. Chem. 274:20587-20596).

Kynurenine aminotransferase catalyzes the irreversible transamination of the L-tryptophan metabolite L-kynurenine to form kynurenic acid. The enzyme may also catalyze the reversible transamination reaction between L-2-aminoadipate and 2-oxoglutarate to produce 2-oxoadipate and L-glutamate. Kynurenic acid is a putative modulator of glutamatergic neurotransmission; thus a deficiency in kynurenine aminotransferase may be associated with pleotrophic effects (Buchli, R. et al. (1995) J. Biol. Chem. 270:29330-29335).

Catechol-*O*-methyltransferase

Catechol-*O*-methyltransferase (COMT) catalyzes the transfer of the methyl group of *S*-adenosyl-L-methionine (AdoMet; SAM) donor to one of the hydroxyl groups of the catechol substrate (e.g., L-dopa, dopamine, or DBA). Methylation of the 3'-hydroxyl group is favored over methylation of the 4'-hydroxyl group and the membrane bound isoform of COMT is more regiospecific than the soluble form. Translation of the soluble form of the enzyme results from utilization of an internal start codon in a full-length mRNA (1.5 kb) or from the translation of a shorter mRNA (1.3 kb), transcribed from an internal promoter. The proposed S_N2-like methylation reaction requires Mg^{++} and is inhibited by Ca^{++}. The binding of the donor and substrate to COMT occurs sequentially. AdoMet first binds COMT in a Mg^{++}-independent manner, followed by the binding of Mg^{++} and the binding of the catechol substrate.

The amount of COMT in tissues is relatively high compared to the amount of activity normally required, thus inhibition is problematic. Nonetheless, inhibitors have been developed for *in vitro* use (e.g., gallates, tropolone, U-0521, and 3',4'-dihydroxy-2-methyl-propiophetropolone) and for

clinical use (e.g., nitrocatechol-based compounds and tolcapone). Administration of these inhibitors results in the increased half-life of L-dopa and the consequent formation of dopamine. Inhibition of COMT is also likely to increase the half-life of various other catechol-structure compounds, including but not limited to epinephrine/norepinephrine, isoprenaline, rimiterol, dobutamine, fenoldopam, apomorphine, and α-methyldopa. A deficiency in norepinephrine has been linked to clinical depression, hence the use of COMT inhibitors could be usefull in the treatment of depression. COMT inhibitors are generally well tolerated with minimal side effects and are ultimately metabolized in the liver with only minor accumulation of metabolites in the body (Männistö, P.T. and S. Kaakkola (1999) Pharmacol. Rev. 51:593-628).

Copper-zinc superoxide dismutases

Copper-zinc superoxide dismutases are compact homodimeric metalloenzymes involved in cellular defenses against oxidative damage. The enzymes contain one atom of zinc and one atom of copper per subunit and catalyze the dismutation of superoxide anions into O_2 and H_2O_2. The rate of dismutation is diffusion-limited and consequently enhanced by the presence of favorable electrostatic interactions between the substrate and enzyme active site. Examples of this class of enzyme have been identified in the cytoplasm of all the eukaryotic cells as well as in the periplasm of several bacterial species. Copper-zinc superoxide dismutases are robust enzymes that are highly resistant to proteolytic digestion and denaturing by urea and SDS. In addition to the compact structure of the enzymes, the presence of the metal ions and intrasubunit disulfide bonds is believed to be responsible for enzyme stability. The enzymes undergo reversible denaturation at temperatures as high as 70 °C (Battistoni, A. et al. (1998) J. Biol. Chem. 273:5655-5661).

Overexpression of superoxide dismutase has been implicated in enhancing freezing tolerance of transgenic alfalfa as well as providing resistance to environmental toxins such as the diphenyl ether herbicide, acifluorfen (McKersie, B.D. et al. (1993) Plant Physiol. 103:1155-1163). In addition, yeast cells become more resistant to freeze-thaw damage following exposure to hydrogen peroxide which causes the yeast cells to adapt to further peroxide stress by upregulating expression of superoxide dismutases. In this study, mutations to yeast superoxide dismutase genes had a more detrimental effect on freeze-thaw resistance than mutations which affected the regulation of glutathione metabolism, long suspected of being important in determining an organism's survival through the process of cryopreservation (Jong-In Park, J.-I. et al. (1998) J. Biol. Chem. 273:22921-22928).

Expression of superoxide dismutase is also associated with *Mycobacterium tuberculosis*, the organism that causes tuberculosis. Superoxide dismutase is one of the ten major proteins secreted by *M. tuberculosis* and its expression is upregulated approximately 5-fold in response to oxidative stress. *M. tuberculosis* expresses almost two orders of magnitude more superoxide dismutase than the

nonpathogenic mycobacterium *M. smegmatis*, and secretes a much higher proportion of the expressed enzyme. The result is the secretion of ~350-fold more enzyme by *M. tuberculosis* than *M. smegmatis*, providing substantial resistance to oxidative stress (Harth, G. and M.A. Horwitz (1999) J. Biol. Chem. 274:4281-4292).

The reduced expression of copper-zinc superoxide dismutases, as well as other enzymes with anti-oxidant capabilities, has been implicated in the early stages of cancer. The expression of copper-zinc superoxide dismutases is reduced in prostatic intraepithelial neoplasia and prostate carcinomas, (Bostwick, D.G. (2000) Cancer 89:123-134).

Phosphoesterases

Phosphotriesterases (PTE, paraoxonases) are enzymes that hydrolyze toxic organophosphorus compounds and have been isolated from a variety of tissues. Phosphotriesterases play a central role in the detoxification of insecticides by mammals. Birds and insects lack PTE, and as a result have reduced tolerance for organophosphorus compounds (Vilanova, E. and M.A. Sogorb (1999) Crit. Rev. Toxicol. 29:21-57). Phosphotriesterase activity varies among individuals and is lower in infants than adults. PTE knockout mice are markedly more sensitive to the organophosphate-based toxins diazoxon and chlorpyrifos oxon (Furlong, C.E., et al. (2000) Neurotoxicology 21:91-100). Phosphotriesterase is also implicated in atherosclerosis and diseases involving lipoprotein metabolism.

Glycerophosphoryl diester phosphodiesterase (also known as glycerophosphodiester phosphodiesterase) is a phosphodiesterase which hydrolyzes deacetylated phospholipid glycerophosphodiesters to produce sn-glycerol-3-phosphate and an alcohol. Glycerophosphocholine, glycerophosphoethanolamine, glycerophosphoglycerol, and glycerophosphoinositol are examples of substrates for glycerophosphoryl diester phosphodiesterases. A glycerophosphoryl diester phosphodiesterase from *E. coli* has broad specificity for glycerophosphodiester substrates (Larson, T.J. et al. (1983) J. Biol. Chem. 248:5428-5432).

Cyclic nucleotide phosphodiesterases (PDEs) are crucial enzymes in the regulation of the cyclic nucleotides cAMP and cGMP. cAMP and cGMP function as intracellular second messengers to transduce a variety of extracellular signals including hormones, light, and neurotransmitters. PDEs degrade cyclic nucleotides to their corresponding monophosphates, thereby regulating the intracellular concentrations of cyclic nucleotides and their effects on signal transduction. Due to their roles as regulators of signal transduction, PDEs have been extensively studied as chemotherapeutic targets (Perry, M.J. and G.A. Higgs (1998) Curr. Opin. Chem. Biol. 2:472-481; Torphy, J.T. (1998) Am. J. Resp. Crit. Care Med. 157:351-370).

Families of mammalian PDEs have been classified based on their substrate specificity and affinity, sensitivity to cofactors, and sensitivity to inhibitory agents (Beavo, J.A. (1995) Physiol. Rev.

75:725-748; Conti, M. et al. (1995) Endocrine Rev. 16:370-389). Several of these families contain distinct genes, many of which are expressed in different tissues as splice variants. Within PDE families, there are multiple isozymes and multiple splice variants of these isozymes (Conti, M. and S.-L.C. Jin (1999) Prog. Nucleic Acid Res. Mol. Biol. 63:1-38). The existence of multiple PDE families, isozymes, and splice variants is an indication of the variety and complexity of the regulatory pathways involving cyclic nucleotides (Houslay, M.D. and G. Milligan (1997) Trends Biochem. Sci. 22:217-224).

Type 1 PDEs (PDE1s) are Ca^{2+}/calmodulin-dependent and appear to be encoded by at least three different genes, each having at least two different splice variants (Kakkar, R. et al. (1999) Cell Mol. Life Sci. 55:1164-1186). PDE1s have been found in the lung, heart, and brain. Some PDE1 isozymes are regulated *in vitro* by phosphorylation/dephosphorylation. Phosphorylation of these PDE1 isozymes decreases the affinity of the enzyme for calmodulin, decreases PDE activity, and increases steady state levels of cAMP (Kakkar et al., *supra*). PDE1s may provide useful therapeutic targets for disorders of the central nervous system and the cardiovascular and immune systems, due to the involvement of PDE1s in both cyclic nucleotide and calcium signaling (Perry and Higgs, *supra*).

PDE2s are cGMP-stimulated PDEs that have been found in the cerebellum, neocortex, heart, kidney, lung, pulmonary artery, and skeletal muscle (Sadhu, K. et al. (1999) J. Histochem. Cytochem. 47:895-906). PDE2s are thought to mediate the effects of cAMP on catecholamine secretion, participate in the regulation of aldosterone (Beavo, *supra*), and play a role in olfactory signal transduction (Juilfs, D.M. et al. (1997) Proc. Natl. Acad. Sci. USA 94:3388-3395).

PDE3s have high affinity for both cGMP and cAMP, and so these cyclic nucleotides act as competitive substrates for PDE3s. PDE3s play roles in stimulating myocardial contractility, inhibiting platelet aggregation, relaxing vascular and airway smooth muscle, inhibiting proliferation of T-lymphocytes and cultured vascular smooth muscle cells, and regulating catecholamine-induced release of free fatty acids from adipose tissue. The PDE3 family of phosphodiesterases are sensitive to specific inhibitors such as cilostamide, enoximone, and lixazinone. Isozymes of PDE3 can be regulated by cAMP-dependent protein kinase, or by insulin-dependent kinases (Degerman, E. et al. (1997) J. Biol. Chem. 272:6823-6826).

PDE4s are specific for cAMP; are localized to airway smooth muscle, the vascular endothelium, and all inflammatory cells; and can be activated by cAMP-dependent phosphorylation. Since elevation of cAMP levels can lead to suppression of inflammatory cell activation and to relaxation of bronchial smooth muscle, PDE4s have been studied extensively as possible targets for novel anti-inflammatory agents, with special emphasis placed on the discovery of asthma treatments. PDE4 inhibitors are currently undergoing clinical trials as treatments for asthma, chronic obstructive pulmonary disease, and atopic eczema. All four known isozymes of PDE4 are susceptible to the

inhibitor rolipram, a compound which has been shown to improve behavioral memory in mice (Barad, M. et al. (1998) Proc. Natl. Acad. Sci. USA 95:15020-15025). PDE4 inhibitors have also been studied as possible therapeutic agents against acute lung injury, endotoxemia, rheumatoid arthritis, multiple sclerosis, and various neurological and gastrointestinal indications (Doherty, A.M. (1999) Curr. Opin. Chem. Biol. 3:466-473).

PDE5 is highly selective for cGMP as a substrate (Turko, I.V. et al. (1998) Biochemistry 37:4200-4205), and has two allosteric cGMP-specific binding sites (McAllister-Lucas, L.M. et al. (1995) J. Biol. Chem. 270:30671-30679). Binding of cGMP to these allosteric binding sites seems to be important for phosphorylation of PDE5 by cGMP-dependent protein kinase rather than for direct regulation of catalytic activity. High levels of PDE5 are found in vascular smooth muscle, platelets, lung, and kidney. The inhibitor zaprinast is effective against PDE5 and PDE1s. Modification of zaprinast to provide specificity against PDE5 has resulted in sildenafil (VIAGRA; Pfizer, Inc., New York NY), a treatment for male erectile dysfunction (Terrett, N. et al. (1996) Bioorg. Med. Chem. Lett. 6:1819-1824). Inhibitors of PDE5 are currently being studied as agents for cardiovascular therapy (Perry and Higgs, *supra*).

PDE6s, the photoreceptor cyclic nucleotide phosphodiesterases, are crucial components of the phototransduction cascade. In association with the G-protein transducin, PDE6s hydrolyze cGMP to regulate cGMP-gated cation channels in photoreceptor membranes. In addition to the cGMP-binding active site, PDE6s also have two high-affinity cGMP-binding sites which are thought to play a regulatory role in PDE6 function (Artemyev, N.O. et al. (1998) Methods 14:93-104). Defects in PDE6s have been associated with retinal disease. Retinal degeneration in the rd mouse (Yan, W. et al. (1998) Invest. Opthalmol. Vis. Sci. 39:2529-2536), autosomal recessive retinitis pigmentosa in humans (Danciger, M. et al. (1995) Genomics 30:1-7), and rod/cone dysplasia 1 in Irish Setter dogs (Suber, M.L. et al. (1993) Proc. Natl. Acad. Sci. USA 90:3968-3972) have been attributed to mutations in the PDE6B gene.

The PDE7 family of PDEs consists of only one known member having multiple splice variants (Bloom, T.J. and J.A. Beavo (1996) Proc. Natl. Acad. Sci. USA 93:14188-14192). PDE7s are cAMP specific, but little else is known about their physiological function. Although mRNAs encoding PDE7s are found in skeletal muscle, heart, brain, lung, kidney, and pancreas, expression of PDE7 proteins is restricted to specific tissue types (Han, P. et al. (1997) J. Biol. Chem. 272:16152-16157; Perry and Higgs, *supra*). PDE7s are very closely related to the PDE4 family; however, PDE7s are not inhibited by rolipram, a specific inhibitor of PDE4s (Beavo, *supra*).

PDE8s are cAMP specific, and are closely related to the PDE4 family. PDE8s are expressed in thyroid gland, testis, eye, liver, skeletal muscle, heart, kidney, ovary, and brain. The cAMP-hydrolyzing activity of PDE8s is not inhibited by the PDE inhibitors rolipram, vinpocetine, milrinone,

IBMX (3-isobutyl-1-methylxanthine), or zaprinast, but PDE8s are inhibited by dipyridamole (Fisher, D.A. et al. (1998) Biochem. Biophys. Res. Commun. 246:570-577; Hayashi, M. et al. (1998) Biochem. Biophys. Res. Commun. 250:751-756; Soderling, S.H. et al. (1998) Proc. Natl. Acad. Sci. USA 95:8991-8996).

PDE9s are cGMP specific and most closely resemble the PDE8 family of PDEs. PDE9s are expressed in kidney, liver, lung, brain, spleen, and small intestine. PDE9s are not inhibited by sildenafil (VIAGRA; Pfizer, Inc., New York NY), rolipram, vinpocetine, dipyridamole, or IBMX (3-isobutyl-1-methylxanthine), but they are sensitive to the PDE5 inhibitor zaprinast (Fisher, D.A. et al. (1998) J. Biol. Chem. 273:15559-15564; Soderling, S.H. et al. (1998) J. Biol. Chem. 273:15553-15558).

PDE10s are dual-substrate PDEs, hydrolyzing both cAMP and cGMP. PDE10s are expressed in brain, thyroid, and testis. (Soderling, S.H. et al. (1999) Proc. Natl. Acad. Sci. USA 96:7071-7076; Fujishige, K. et al. (1999) J. Biol. Chem. 274:18438-18445; Loughney, K. et al (1999) Gene 234:109-117).

PDEs are composed of a catalytic domain of about 270-300 amino acids, an N-terminal regulatory domain responsible for binding cofactors, and, in some cases, a hydrophilic C-terminal domain of unknown function (Conti and Jin, *supra*). A conserved, putative zinc-binding motif has been identified in the catalytic domain of all PDEs. N-terminal regulatory domains include non-catalytic cGMP-binding domains in PDE2s, PDE5s, and PDE6s; calmodulin-binding domains in PDE1s; and domains containing phosphorylation sites in PDE3s and PDE4s. In PDE5, the N-terminal cGMP-binding domain spans about 380 amino acid residues and comprises tandem repeats of a conserved sequence motif (McAllister-Lucas, L.M. et al. (1993) J. Biol. Chem. 268:22863-22873). The NKXnD motif has been shown by mutagenesis to be important for cGMP binding (Turko, I.V. et al. (1996) J. Biol. Chem. 271:22240-22244). PDE families display approximately 30% amino acid identity within the catalytic domain; however, isozymes within the same family typically display about 85-95% identity in this region (e.g. PDE4A vs PDE4B). Furthermore, within a family there is extensive similarity (>60%) outside the catalytic domain; while across families, there is little or no sequence similarity outside this domain.

Many of the constituent functions of immune and inflammatory responses are inhibited by agents that increase intracellular levels of cAMP (Verghese, M.W. et al. (1995) Mol. Pharmacol. 47:1164-1171). A variety of diseases have been attributed to increased PDE activity and associated with decreased levels of cyclic nucleotides. For example, a form of diabetes insipidus in mice has been associated with increased PDE4 activity, an increase in low-K_m cAMP PDE activity has been reported in leukocytes of atopic patients, and PDE3 has been associated with cardiac disease.

Many inhibitors of PDEs have undergone clinical evaluation (Perry and Higgs, *supra*;

Torphy, T.J. (1998) Am. J. Respir. Crit. Care Med. 157:351-370). PDE3 inhibitors are being developed as antithrombotic agents, antihypertensive agents, and as cardiotonic agents useful in the treatment of congestive heart failure. Rolipram, a PDE4 inhibitor, has been used in the treatment of depression, and other PDE4 inhibitors have an anti-inflammatory effect. Rolipram may inhibit HIV-1 replication (Angel, J.B. et al. (1995) AIDS 9:1137-1144). Additionally, rolipram suppresses the production of cytokines such as TNF-a and b and interferon g, and thus is effective against encephalomyelitis. Rolipram may also be effective in treating tardive dyskinesia and multiple sclerosis (Sommer, N. et al. (1995) Nat. Med. 1:244-248; Sasaki, H. et al. (1995) Eur. J. Pharmacol. 282:71-76). Theophylline is a nonspecific PDE inhibitor used in treatment of bronchial asthma and other respiratory diseases. Theophylline is believed to act on airway smooth muscle function and in an anti-inflammatory or immunomodulatory capacity (Banner, K.H. and C.P. Page (1995) Eur. Respir. J. 8:996-1000). Pentoxifylline is another nonspecific PDE inhibitor used in the treatment of intermittent claudication and diabetes-induced peripheral vascular disease. Pentoxifylline is also known to block TNF-a production and may inhibit HIV-1 replication (Angel et al., *supra*).

PDEs have been reported to affect cellular proliferation of a variety of cell types (Conti et al. (1995) Endocrine Rev. 16:370-389) and have been implicated in various cancers. Growth of prostate carcinoma cell lines DU145 and LNCaP was inhibited by delivery of cAMP derivatives and PDE inhibitors (Bang, Y.J. et al. (1994) Proc. Natl. Acad. Sci. USA 91:5330-5334). These cells also showed a permanent conversion in phenotype from epithelial to neuronal morphology. It has also been suggested that PDE inhibitors can regulate mesangial cell proliferation (Matousovic, K. et al. (1995) J. Clin. Invest. 96:401-410) and lymphocyte proliferation (Joulain, C. et al. (1995) J. Lipid Mediat. Cell Signal. 11:63-79). One cancer treatment involves intracellular delivery of PDEs to particular cellular compartments of tumors, resulting in cell death (Deonarain, M.P. and A.A. Epenetos (1994) Br. J. Cancer 70:786-794).

Members of the UDP glucuronyltransferase family (UGTs) catalyze the transfer of a glucuronic acid group from the cofactor uridine diphosphate-glucuronic acid (UDP-glucuronic acid) to a substrate. The transfer is generally to a nucleophilic heteroatom (O, N, or S). Substrates include xenobiotics which have been functionalized by Phase I reactions, as well as endogenous compounds such as bilirubin, steroid hormones, and thyroid hormones. Products of glucuronidation are excreted in urine if the molecular weight of the substrate is less than about 250 g/mol, whereas larger glucuronidated substrates are excreted in bile.

UGTs are located in the microsomes of liver, kidney, intestine, skin, brain, spleen, and nasal mucosa, where they are on the same side of the endoplasmic reticulum membrane as cytochrome P450 enzymes and flavin-containing monooxygenases. UGTs have a C-terminal membrane-spanning domain which anchors them in the endoplasmic reticulum membrane, and a conserved signature

domain of about 50 amino acid residues in their C terminal section (PROSITE PDOC00359 UDP-glycosyltransferase signature).

UGTs involved in drug metabolism are encoded by two gene families, UGT1 and UGT2. Members of the UGT1 family result from alternative splicing of a single gene locus, which has a variable substrate binding domain and constant region involved in cofactor binding and membrane insertion. Members of the UGT2 family are encoded by separate gene loci, and are divided into two families, UGT2A and UGT2B. The 2A subfamily is expressed in olfactory epithelium, and the 2B subfamily is expressed in liver microsomes. Mutations in UGT genes are associated with hyperbilirubinemia (OMIM #143500 Hyperbilirubinemia I); Crigler-Najjar syndrome, characterized by intense hyperbilirubinemia from birth (OMIM #218800 Crigler-Najjar syndrome); and a milder form of hyperbilirubinemia termed Gilbert's disease (OMIM #191740 UGT1).

Thioesterases

Two soluble thioesterases involved in fatty acid biosynthesis have been isolated from mammalian tissues, one which is active only toward long-chain fatty-acyl thioesters and one which is active toward thioesters with a wide range of fatty-acyl chain-lengths. These thioesterases catalyze the chain-terminating step in the *de novo* biosynthesis of fatty acids. Chain termination involves the hydrolysis of the thioester bond which links the fatty acyl chain to the 4'-phosphopantetheine prosthetic group of the acyl carrier protein (ACP) subunit of the fatty acid synthase (Smith, S. (1981a) Methods Enzymol. 71:181-188; Smith, S. (1981b) Methods Enzymol. 71:188-200).

E. coli contains two soluble thioesterases, thioesterase I which is active only toward long-chain acyl thioesters, and thioesterase II (TEII) which has a broad chain-length specificity (Naggert, J. et al. (1991) J. Biol. Chem. 266:11044-11050). *E. coli* TEII does not exhibit sequence similarity with either of the two types of mammalian thioesterases which function as chain-terminating enzymes in *de novo* fatty acid biosynthesis. Unlike the mammalian thioesterases, *E. coli* TEII lacks the characteristic serine active site gly-X-ser-X-gly sequence motif and is not inactivated by the serine modifying agent diisopropyl fluorophosphate. However, modification of histidine 58 by iodoacetamide and diethylpyrocarbonate abolished TEII activity. Overexpression of TEII did not alter fatty acid content in *E. coli*, which suggests that it does not function as a chain-terminating enzyme in fatty acid biosynthesis (Naggert et al., *supra*). For that reason, Naggert et al. (*supra*) proposed that the physiological substrates for *E. coli* TEII may be coenzyme A (CoA)-fatty acid esters instead of ACP-phosphopanthetheine-fatty acid esters.

Carboxylesterases

Mammalian carboxylesterases are a multigene family expressed in a variety of tissues and cell types. Acetylcholinesterase, butyrylcholinesterase, and carboxylesterase are grouped into the serine superfamily of esterases (B-esterases). Other carboxylesterases include thyroglobulin,

thrombin, Factor IX, gliotactin, and plasminogen. Carboxylesterases catalyze the hydrolysis of ester- and amide- groups from molecules and are involved in detoxification of drugs, environmental toxins, and carcinogens. Substrates for carboxylesterases include short- and long-chain acyl-glycerols, acylcarnitine, carbonates, dipivefrin hydrochloride, cocaine, salicylates, capsaicin, palmitoyl-coenzyme A, imidapril, haloperidol, pyrrolizidine alkaloids, steroids, p-nitrophenyl acetate, malathion, butanilicaine, and isocarboxazide. Carboxylesterases are also important for the conversion of prodrugs to free acids, which may be the active form of the drug (e.g., lovastatin, used to lower blood cholesterol) (reviewed in Satoh, T. and Hosokawa, M. (1998) Annu. Rev. Pharmacol. Toxicol.38:257-288). Neuroligins are a class of molecules that (i) have N-terminal signal sequences, (ii) resemble cell-surface receptors, (iii) contain carboxylesterase domains, (iv) are highly expressed in the brain, and (v) bind to neurexins in a calcium-dependent manner. Despite the homology to carboxylesterases, neuroligins lack the active site serine residue, implying a role in substrate binding rather than catalysis (Ichtchenko, K. et al. (1996) J. Biol. Chem. 271:2676-2682).

Squalene epoxidase

Squalene epoxidase (squalene monooxygenase, SE) is a microsomal membrane-bound, FAD-dependent oxidoreductase that catalyzes the first oxygenation step in the sterol biosynthetic pathway of eukaryotic cells. Cholesterol is an essential structural component of cytoplasmic membranes acquired via the LDL receptor-mediated pathway or the biosynthetic pathway. SE converts squalene to 2,3(*S*)-oxidosqualene, which is then converted to lanosterol and then cholesterol.

High serum cholesterol levels result in the formation of atherosclerotic plaques in the arteries of higher organisms. This deposition of highly insoluble lipid material onto the walls of essential blood vessels results in decreased blood flow and potential necrosis. HMG-CoA reductase is responsible for the first committed step in cholesterol biosynthesis, conversion of 3-hydroxyl-3-methyl-glutaryl CoA (HMG-CoA) to mevalonate. HMG-CoA is the target of a number of pharmaceutical compounds designed to lower plasma cholesterol levels, but inhibition of MHG-CoA also results in the reduced synthesis of non-sterol intermediates required for other biochemical pathways. Since SE catalyzes a rate-limiting reaction that occurs later in the sterol synthesis pathway with cholesterol as the only end product, SE is a better ideal target for the design of anti-hyperlipidemic drugs (Nakamura, Y. et al. (1996) 271:8053-8056).

Epoxide hydrolases

Epoxide hydrolases catalyze the addition of water to epoxide-containing compounds, thereby hydrolyzing epoxides to their corresponding 1,2-diols. They are related to bacterial haloalkane dehalogenases and show sequence similarity to other members of the α/β hydrolase fold family of enzymes. This family of enzymes is important for the detoxification of xenobiotic epoxide compounds which are often highly electrophilic and destructive when introduced. Examples of

epoxide hydrolase reactions include the hydrolysis of some leukotoxin to leukotoxin diol, and isoleukotoxin to isoleukotoxin diol. Leukotoxins alter membrane permeability and ion transport and cause inflammatory responses. In addition, epoxide carcinogens are produced by cytochrome P450 as intermediates in the detoxification of drugs and environmental toxins. Epoxide hydrolases possess a catalytic triad composed of Asp, Asp, and His (Arand, M. et al. (1996) J. Biol. Chem. 271:4223-4229; Rink, R. et al. (1997) J. Biol. Chem. 272:14650-14657; Argiriadi, M.A. et al. (2000) J. Biol. Chem. 275:15265-15270).

Enzymes involved in tyrosine catalysis

The degradation of the amino acid tyrosine, to either succinate and pyruvate or fumarate and acetoacetate, requires a large number of enzymes and generates a large number of intermediate compounds. In addition, many xenobiotic compounds may be metabolized using one or more reactions that are part of the tyrosine catabolic pathway. Enzymes involved in the degradation of tyrosine to succinate and pyruvate (e.g., in *Arthrobacter* species) include 4-hydroxyphenylpyruvate oxidase, 4-hydroxyphenylacetate 3-hydroxylase, 3,4-dihydroxyphenylacetate 2,3-dioxygenase, 5-carboxymethyl-2-hydroxymuconic semialdehyde dehydrogenase, *trans,cis*-5-carboxymethyl-2-hydroxymuconate isomerase, homoprotocatechuate isomerase/decarboxylase, *cis*-2-oxohept-3-ene-1,7-dioate hydratase, 2,4-dihydroxyhept-*trans*-2-ene-1,7-dioate aldolase, and succinic semialdehyde dehydrogenase. Enzymes involved in the degradation of tyrosine to fumarate and acetoacetate (e.g., in *Pseudomonas* species) include 4-hydroxyphenylpyruvate dioxygenase, homogentisate 1,2-dioxygenase, maleylacetoacetate isomerase, fumarylacetoacetase and 4-hydroxyphenylacetate. Additional enzymes associated with tyrosine metabolism in different organisms include 4-chlorophenylacetate-3,4-dioxygenase, aromatic aminotransferase, 5-oxopent-3-ene-1,2,5-tricarboxylate decarboxylase, 2-oxo-hept-3-ene-1,7-dioate hydratase, and 5-carboxymethyl-2-hydroxymuconate isomerase (Ellis, L.B.M. et al. (1999) Nucleic Acids Res. 27:373-376; Wackett, L.P. and Ellis, L.B.M. (1996) J. Microbiol. Meth. 25:91-93; and Schmidt, M. (1996) Amer. Soc. Microbiol. News 62:102).

In humans, acquired or inherited genetic defects in enzymes of the tyrosine degradation pathway may result in hereditary tyrosinemia. One form of this disease, hereditary tyrosinemia 1 (HT1) is caused by a deficiency in the enzyme fumarylacetoacetate hydrolase, the last enzyme in the pathway in organisms that metabolize tyrosine to fumarate and acetoacetate. HT1 is characterized by progressive liver damage beginning at infancy, and increased risk for liver cancer (Endo, F. et al. (1997) J. Biol. Chem. 272:24426-24432).

Parkinson's Disease

Parkinson's disease (PD) is a neurodegenerative disorder characterized by the progressive degeneration of the dopaminergic nigrostriatal pathway, and the presence of Lewy bodies. Genetic

linkages for the Parkin gene to chromosome 6q25.2-27, for PARK3 to chromosome 2p (West, A. B. (2001) Eur. J. Hum. Genet. 9:659-666), and for PARK6 to chromosome 1p35-p36 have been identified (Valente, E. M. et al. (2002) Ann. Neurol. 51:14-18). Clinical disorders classified as parkinsonism include PD, dementia with Lewy bodies (DLB), progressive supranuclear palsy (PSP), and essential tremor. Several neurodegenerative diseases share pathogenic mechanisms involving tau or synuclein aggregation. These disorders include Alzheimer's disease, and Pick's disease as well as PD and progressive supranuclear palsy (Hardy, J. (2001) J. Alzheimers Dis. 3:109-116). Several genetically distinct forms of PD can be caused by mutations in single genes. Genes for monogenically inherited forms of Parkinson's disease have been mapped and/or cloned. In some families with autosomal dominant inheritance and typical Lewy-body pathology, mutations have been identified in the gene for alpha-synuclein. Aggregation of this protein in Lewy-bodies may be a crucial step in the molecular pathogenesis of familial and sporadic PD.

Parkin-mutations appear to be a common cause of PD in patients with very early onset. Mutations in the parkin gene of early-onset PD are autosomal recessive mutations in which nigral degeneration is not accompanied by Lewy-body formation. Parkin has been implicated in the cellular protein degradation pathways, as it has been shown that it functions as a ubiquitin ligase. A mutation in the gene for ubiquitin C-terminal hydrolase L1 in this pathway has been identified in another small family with PD. Other loci have been mapped to chromosome 2p and 4p, respectively, in families with dominantly inherited PD. These early-onset forms differ from the common sporadic form of PD. It is widely believed that a combination of interacting genetic and environmental causes may be responsible in the majority of PD-cases (Gasser, T. (2001) J. Neurol. 2001 248:833-840).

Alzheimer's Disease

Characterization of region-specific gene expression in the human brain provides a context and background for molecular neurobiology on a variety of neurological disorders. For example, Alzheimer's disease (AD) is a progressive, neurodestructive process of the human neocortex, characterized by the deterioration of memory and higher cognitive function. A progressive and irreversible brain disorder, AD is characterized by three major pathogenic episodes involving (a) an aberrant processing and deposition of beta-amyloid precursor protein (betaAPP) to form neurotoxic beta-amyloid (betaA) peptides and an aggregated insoluble polymer of betaA that forms the senile plaque, (b) the establishment of intraneuronal neuritic tau pathology yielding widespread deposits of agyrophilic neurofibrillary tangles (NFT) and (c) the initiation and proliferation of a brain-specific inflammatory response. These three seemingly disperse attributes of AD etiopathogenesis are linked by the fact that proinflammatory microglia, reactive astrocytes and their associated cytokines and chemokines are associated with the biology of the microtubule associated protein tau, betaA speciation and aggregation. Missense mutations in the presenilin genes PS1 and PS2, implicated in

early onset familial AD, cause abnormal betaAPP processing with resultant overproduction of betaA42 and related neurotoxic peptides. Specific betaA fragments such as betaA42 can further potentiate proinflammatory mechanisms. Expression of the inducible oxidoreductase cyclooxygenase-2 and cytosolic phospholipase A2 (cPLA2) is strongly activated during cerebral ischemia and trauma, epilepsy and AD, indicating the induction of proinflammatory gene pathways as a response to brain injury. Neurotoxic metals such as aluminum and zinc, both implicated in AD etiopathogenesis, and arachidonic acid, a major metabolite of brain cPLA2 activity, each polymerize hyperphosphorylated tau to form NFT-like bundles. Studies have identified a reduced risk for AD in patients aged over 70 years who were previously treated with non-steroidal anti-inflammatory drugs for non-CNS afflictions that include arthritis. (For a review of the interrelationships between the mechanisms of PS1, PS2 and betaAPP gene expression, tau and betaA deposition and the induction, regulation and proliferation in AD of the neuroinflammatory response, see Lukiw, W.J, and Bazan, N.G. (2000) Neurochem. Res. 2000 25:1173-1184).

Expression Profiling

Microarrays are analytical tools used in bioanalysis. A microarray has a plurality of molecules spatially distributed over, and stably associated with, the surface of a solid support. Microarrays of polypeptides, polynucleotides, and/or antibodies have been developed and find use in a variety of applications, such as gene sequencing, monitoring gene expression, gene mapping, bacterial identification, drug discovery, and combinatorial chemistry.

One area in particular in which microarrays find use is in gene expression analysis. Array technology can provide a simple way to explore the expression of a single polymorphic gene or the expression profile of a large number of related or unrelated genes. When the expression of a single gene is examined, arrays are employed to detect the expression of a specific gene or its variants. When an expression profile is examined, arrays provide a platform for identifying genes that are tissue specific, are affected by a substance being tested in a toxicology assay, are part of a signaling cascade, carry out housekeeping functions, or are specifically related to a particular genetic predisposition, condition, disease, or disorder.

While soft tissue sarcomas are relatively rare, more than 50% of new patients diagnosed with the disease will die from it. The molecular pathways leading to the development of sarcomas are relatively unknown, due to the rarity of the disease and variation in pathology. Colon cancer evolves through a multi-step process whereby pre-malignant colonocytes undergo a relatively defined sequence of events leading to tumor formation. Several factors participate in the process of tumor progression and malignant transformation including genetic factors, mutations, and selection.

To understand the nature of gene alterations in colorectal cancer, a number of studies have focused on the inherited syndromes. Familial adenomatous polyposis (FAP), is caused by mutations

in the adenomatous polyposis coli gene (APC), resulting in truncated or inactive forms of the protein. This tumor suppressor gene has been mapped to chromosome 5q. Hereditary nonpolyposis colorectal cancer (HNPCC) is caused by mutations in mis-match repair genes. Although hereditary colon cancer syndromes occur in a small percentage of the population and most colorectal cancers are considered sporadic, knowledge from studies of the hereditary syndromes can be generally applied. For instance, somatic mutations in APC occur in at least 80% of sporadic colon tumors. APC mutations are thought to be the initiating event in the disease. Other mutations occur subsequently. Approximately 50% of colorectal cancers contain activating mutations in ras, while 85% contain inactivating mutations in p53. Changes in all of these genes lead to gene expression changes in colon cancer.

Ovarian cancer is the leading cause of death from a gynecologic cancer. The majority of ovarian cancers are derived from epithelial cells, and 70% of patients with epithelial ovarian cancers present with late-stage disease. As a result, the long-term survival rates for this disease is very low. Identification of early-stage markers for ovarian cancer would significantly increase the survival rate. Genetic variations involved in ovarian cancer development include mutation of p53 and microsatellite instability. Gene expression patterns likely vary when normal ovary is compared to ovarian tumors.

There are more than 180,000 new cases of breast cancer diagnosed each year, and the mortality rate for breast cancer approaches 10% of all deaths in females between the ages of 45-54 (Gish, K. (1999) AWIS Magazine 28:7-10). However the survival rate based on early diagnosis of localized breast cancer is extremely high (97%), compared with the advanced stage of the disease in which the tumor has spread beyond the breast (22%). Current procedures for clinical breast examination are lacking in sensitivity and specificity, and efforts are underway to develop comprehensive gene expression profiles for breast cancer that may be used in conjunction with conventional screening methods to improve diagnosis and prognosis of this disease (Perou, C.M. et al. (2000) Nature 406:747-752).

Mutations in two genes, BRCA1 and BRCA2, are known to greatly predispose a woman to breast cancer and may be passed on from parents to children (Gish, supra). However, this type of hereditary breast cancer accounts for only about 5% to 9% of breast cancers, while the vast majority of breast cancer is due to non-inherited mutations that occur in breast epithelial cells.

The relationship between expression of epidermal growth factor (EGF) and its receptor, EGFR, to human mammary carcinoma has been particularly well studied. (See Khazaie, K. et al. (1993) Cancer and Metastasis Rev. 12:255-274, and references cited therein for a review of this area.) Overexpression of EGFR, particularly coupled with down-regulation of the estrogen receptor, is a marker of poor prognosis in breast cancer patients. In addition, EGFR expression in breast tumor metastases is frequently elevated relative to the primary tumor, suggesting that EGFR is involved in tumor progression and metastasis. This is supported by accumulating evidence that EGF has effects

on cell functions related to metastatic potential, such as cell motility, chemotaxis, secretion and differentiation. Changes in expression of other members of the erbB receptor family, of which EGFR is one, have also been implicated in breast cancer. The abundance of erbB receptors, such as HER-2/neu, HER-3, and HER-4, and their ligands in breast cancer points to their functional importance in the pathogenesis of the disease, and may therefore provide targets for therapy of the disease (Bacus, S. S. et al. (1994) Am. J. Clin. Pathol. 102:S13-S24). Other known markers of breast cancer include a human secreted frizzled protein mRNA that is downregulated in breast tumors; the matrix G1a protein which is overexpressed is human breast carcinoma cells; Drg1 or RTP, a gene whose expression is diminished in colon, breast, and prostate tumors; maspin, a tumor suppressor gene downregulated in invasive breast carcinomas; and CaN19, a member of the S100 protein family, all of which are down regulated in mammary carcinoma cells relative to normal mammary epithelial cells (Zhou, Z. et al. (1998) Int. J. Cancer 78:95-99; Chen, L. et al. (1990) Oncogene 5:1391-1395; Ulrix, W. et al (1999) FEBS Lett 455:23-26; Sager, R. et al. (1996) Curr. Top. Microbiol. Immunol. 213:51-64; and Lee, S. W. et al. (1992) Proc. Natl. Acad. Sci. USA 89:2504-2508).

Cell lines derived from human mammary epithelial cells at various stages of breast cancer provide a useful model to study the process of malignant transformation and tumor progression as it has been shown that these cell lines retain many of the properties of their parental tumors for lengthy culture periods (Wistuba, I.I. et al. (1998) Clin. Cancer Res. 4:2931-2938). Such a model is particularly useful for comparing phenotypic and molecular characteristics of human mammary epithelial cells at various stages of malignant transformation.

Lung cancer is the leading cause of cancer death for men and the second leading cause of cancer death for women in the U.S. Lung cancers are divided into four histopathologically distinct groups. Three groups (squamous cell carcinoma, adenocarcinoma, and large cell carcinoma) are classified as non-small cell lung cancers (NSCLCs). The fourth group of cancers is referred to as small cell lung cancer (SCLC). Deletions on chromosome 3 are common in this disease and are thought to indicate the presence of a tumor suppressor gene in this region. Activating mutations in K-ras are commonly found in lung cancer and are the basis of one of the mouse models for the disease.

There is a need in the art for new compositions, including nucleic acids and proteins, for the diagnosis, prevention, and treatment of autoimmune/inflammatory disorders, infectious disorders, immune deficiencies, disorders of metabolism, reproductive disorders, neurological disorders, cardiovascular disorders, eye disorders, and cell proliferative disorders, including cancer.

SUMMARY OF THE INVENTION

Various embodiments of the invention provide purified polypeptides, enzymes, referred to collectively as 'ENZM' and individually as 'ENZM-1,' 'ENZM-2,' 'ENZM-3,' 'ENZM-4,' 'ENZM-

5,' 'ENZM-6,' 'ENZM-7,' 'ENZM-8,' 'ENZM-9,' 'ENZM-10,' 'ENZM-11,' 'ENZM-12,' 'ENZM-13,' 'ENZM-14,' 'ENZM-15,' 'ENZM-16,' 'ENZM-17,' 'ENZM-18,' 'ENZM-19,' 'ENZM-20,' 'ENZM-21,' 'ENZM-22,' 'ENZM-23,' 'ENZM-24,' 'ENZM-25,' 'ENZM-26,' 'ENZM-27,' 'ENZM-28,' 'ENZM-29,' 'ENZM-30,' 'ENZM-31,' 'ENZM-32,' 'ENZM-33,' 'ENZM-34,' 'ENZM-35,' 'ENZM-36,' 'ENZM-37,' 'ENZM-38,' 'ENZM-39,' 'ENZM-40,' 'ENZM-41,' 'ENZM-42,' 'ENZM-43,' 'ENZM-44,' 'ENZM-45,' 'ENZM-46,' 'ENZM-47,' 'ENZM-48,' 'ENZM-49,' 'ENZM-50,' 'ENZM-51,' 'ENZM-52,' 'ENZM-53,' 'ENZM-54,' and 'ENZM-55' and methods for using these proteins and their encoding polynucleotides for the detection, diagnosis, and treatment of diseases and medical conditions. Embodiments also provide methods for utilizing the purified enzymes and/or their encoding polynucleotides for facilitating the drug discovery process, including determination of efficacy, dosage, toxicity, and pharmacology. Related embodiments provide methods for utilizing the purified enzymes and/or their encoding polynucleotides for investigating the pathogenesis of diseases and medical conditions.

An embodiment provides an isolated polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-55, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-55, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-55, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-55. Another embodiment provides an isolated polypeptide comprising an amino acid sequence of SEQ ID NO:1-55.

Still another embodiment provides an isolated polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-55, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-55, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-55, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-55. In another embodiment, the polynucleotide encodes a polypeptide selected from the group consisting of SEQ ID NO:1-55. In an alternative embodiment, the polynucleotide is selected from the group consisting of SEQ ID NO:56-110.

Still another embodiment provides a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting

of SEQ ID NO:1-55, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-55, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-55, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-55. Another embodiment provides a cell transformed with the recombinant polynucleotide. Yet another embodiment provides a transgenic organism comprising the recombinant polynucleotide.

Another embodiment provides a method for producing a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-55, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-55, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-55, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-55. The method comprises a) culturing a cell under conditions suitable for expression of the polypeptide, wherein said cell is transformed with a recombinant polynucleotide comprising a promoter sequence operably linked to a polynucleotide encoding the polypeptide, and b) recovering the polypeptide so expressed.

Yet another embodiment provides an isolated antibody which specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-55, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-55, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-55, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-55.

Still yet another embodiment provides an isolated polynucleotide selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:56-110, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:56-110, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). In other embodiments, the polynucleotide can comprise at least about 20, 30, 40, 60, 80, or 100 contiguous nucleotides.

Yet another embodiment provides a method for detecting a target polynucleotide in a sample,

said target polynucleotide being selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:56-110, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:56-110, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) hybridizing the sample with a probe comprising at least 20 contiguous nucleotides comprising a sequence complementary to said target polynucleotide in the sample, and which probe specifically hybridizes to said target polynucleotide, under conditions whereby a hybridization complex is formed between said probe and said target polynucleotide or fragments thereof, and b) detecting the presence or absence of said hybridization complex. In a related embodiment, the method can include detecting the amount of the hybridization complex. In still other embodiments, the probe can comprise at least about 20, 30, 40, 60, 80, or 100 contiguous nucleotides.

Still yet another embodiment provides a method for detecting a target polynucleotide in a sample, said target polynucleotide being selected from the group consisting of a) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:56-110, b) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:56-110, c) a polynucleotide complementary to the polynucleotide of a), d) a polynucleotide complementary to the polynucleotide of b), and e) an RNA equivalent of a)-d). The method comprises a) amplifying said target polynucleotide or fragment thereof using polymerase chain reaction amplification, and b) detecting the presence or absence of said amplified target polynucleotide or fragment thereof. In a related embodiment, the method can include detecting the amount of the amplified target polynucleotide or fragment thereof.

Another embodiment provides a composition comprising an effective amount of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-55, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-55, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-55, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-55, and a pharmaceutically acceptable excipient. In one embodiment, the composition can comprise an amino acid sequence selected from the group consisting of SEQ ID NO:1-55. Other embodiments provide a method of treating a disease or condition associated with decreased or abnormal expression of functional ENZM, comprising

administering to a patient in need of such treatment the composition.

Yet another embodiment provides a method for screening a compound for effectiveness as an agonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-55, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-55, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-55, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-55. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting agonist activity in the sample. Another embodiment provides a composition comprising an agonist compound identified by the method and a pharmaceutically acceptable excipient. Yet another embodiment provides a method of treating a disease or condition associated with decreased expression of functional ENZM, comprising administering to a patient in need of such treatment the composition.

Still yet another embodiment provides a method for screening a compound for effectiveness as an antagonist of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-55, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-55, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-55, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-55. The method comprises a) exposing a sample comprising the polypeptide to a compound, and b) detecting antagonist activity in the sample. Another embodiment provides a composition comprising an antagonist compound identified by the method and a pharmaceutically acceptable excipient. Yet another embodiment provides a method of treating a disease or condition associated with overexpression of functional ENZM, comprising administering to a patient in need of such treatment the composition.

Another embodiment provides a method of screening for a compound that specifically binds to a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-55, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-55, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-55, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-55. The method comprises a) combining the

polypeptide with at least one test compound under suitable conditions, and b) detecting binding of the polypeptide to the test compound, thereby identifying a compound that specifically binds to the polypeptide.

Yet another embodiment provides a method of screening for a compound that modulates the activity of a polypeptide selected from the group consisting of a) a polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO:1-55, b) a polypeptide comprising a naturally occurring amino acid sequence at least 90% identical or at least about 90% identical to an amino acid sequence selected from the group consisting of SEQ ID NO:1-55, c) a biologically active fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-55, and d) an immunogenic fragment of a polypeptide having an amino acid sequence selected from the group consisting of SEQ ID NO:1-55. The method comprises a) combining the polypeptide with at least one test compound under conditions permissive for the activity of the polypeptide, b) assessing the activity of the polypeptide in the presence of the test compound, and c) comparing the activity of the polypeptide in the presence of the test compound with the activity of the polypeptide in the absence of the test compound, wherein a change in the activity of the polypeptide in the presence of the test compound is indicative of a compound that modulates the activity of the polypeptide.

Still yet another embodiment provides a method for screening a compound for effectiveness in altering expression of a target polynucleotide, wherein said target polynucleotide comprises a polynucleotide sequence selected from the group consisting of SEQ ID NO:56-110, the method comprising a) exposing a sample comprising the target polynucleotide to a compound, b) detecting altered expression of the target polynucleotide, and c) comparing the expression of the target polynucleotide in the presence of varying amounts of the compound and in the absence of the compound.

Another embodiment provides a method for assessing toxicity of a test compound, said method comprising a) treating a biological sample containing nucleic acids with the test compound; b) hybridizing the nucleic acids of the treated biological sample with a probe comprising at least 20 contiguous nucleotides of a polynucleotide selected from the group consisting of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:56-110, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:56-110, iii) a polynucleotide having a sequence complementary to i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Hybridization occurs under conditions whereby a specific hybridization complex is formed between said probe and a target polynucleotide in the biological sample, said target polynucleotide selected from the group consisting

of i) a polynucleotide comprising a polynucleotide sequence selected from the group consisting of SEQ ID NO:56-110, ii) a polynucleotide comprising a naturally occurring polynucleotide sequence at least 90% identical or at least about 90% identical to a polynucleotide sequence selected from the group consisting of SEQ ID NO:56-110, iii) a polynucleotide complementary to the polynucleotide of i), iv) a polynucleotide complementary to the polynucleotide of ii), and v) an RNA equivalent of i)-iv). Alternatively, the target polynucleotide can comprise a fragment of a polynucleotide selected from the group consisting of i)-v) above; c) quantifying the amount of hybridization complex; and d) comparing the amount of hybridization complex in the treated biological sample with the amount of hybridization complex in an untreated biological sample, wherein a difference in the amount of hybridization complex in the treated biological sample is indicative of toxicity of the test compound.

BRIEF DESCRIPTION OF THE TABLES

Table 1 summarizes the nomenclature for full length polynucleotide and polypeptide embodiments of the invention.

Table 2 shows the GenBank identification number and annotation of the nearest GenBank homolog, and the PROTEOME database identification numbers and annotations of PROTEOME database homologs, for polypeptide embodiments of the invention. The probability scores for the matches between each polypeptide and its homolog(s) are also shown.

Table 3 shows structural features of polypeptide embodiments, including predicted motifs and domains, along with the methods, algorithms, and searchable databases used for analysis of the polypeptides.

Table 4 lists the cDNA and/or genomic DNA fragments which were used to assemble polynucleotide embodiments, along with selected fragments of the polynucleotides.

Table 5 shows representative cDNA libraries for polynucleotide embodiments.

Table 6 provides an appendix which describes the tissues and vectors used for construction of the cDNA libraries shown in Table 5.

Table 7 shows the tools, programs, and algorithms used to analyze polynucleotides and polypeptides, along with applicable descriptions, references, and threshold parameters.

Table 8 shows single nucleotide polymorphisms found in polynucleotide sequences of the invention, along with allele frequencies in different human populations.

DESCRIPTION OF THE INVENTION

Before the present proteins, nucleic acids, and methods are described, it is understood that embodiments of the invention are not limited to the particular machines, instruments, materials, and methods described, as these may vary. It is also to be understood that the terminology used herein is

for the purpose of describing particular embodiments only, and is not intended to limit the scope of the invention.

As used herein and in the appended claims, the singular forms "a," "an," and "the" include plural reference unless the context clearly dictates otherwise. Thus, for example, a reference to "a host cell" includes a plurality of such host cells, and a reference to "an antibody" is a reference to one or more antibodies and equivalents thereof known to those skilled in the art, and so forth.

Unless defined otherwise, all technical and scientific terms used herein have the same meanings as commonly understood by one of ordinary skill in the art to which this invention belongs. Although any machines, materials, and methods similar or equivalent to those described herein can be used to practice or test the present invention, the preferred machines, materials and methods are now described. All publications mentioned herein are cited for the purpose of describing and disclosing the cell lines, protocols, reagents and vectors which are reported in the publications and which might be used in connection with various embodiments of the invention. Nothing herein is to be construed as an admission that the invention is not entitled to antedate such disclosure by virtue of prior invention.

DEFINITIONS

"ENZM" refers to the amino acid sequences of substantially purified ENZM obtained from any species, particularly a mammalian species, including bovine, ovine, porcine, murine, equine, and human, and from any source, whether natural, synthetic, semi-synthetic, or recombinant.

The term "agonist" refers to a molecule which intensifies or mimics the biological activity of ENZM. Agonists may include proteins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of ENZM either by directly interacting with ENZM or by acting on components of the biological pathway in which ENZM participates.

An "allelic variant" is an alternative form of the gene encoding ENZM. Allelic variants may result from at least one mutation in the nucleic acid sequence and may result in altered mRNAs or in polypeptides whose structure or function may or may not be altered. A gene may have none, one, or many allelic variants of its naturally occurring form. Common mutational changes which give rise to allelic variants are generally ascribed to natural deletions, additions, or substitutions of nucleotides. Each of these types of changes may occur alone, or in combination with the others, one or more times in a given sequence.

"Altered" nucleic acid sequences encoding ENZM include those sequences with deletions, insertions, or substitutions of different nucleotides, resulting in a polypeptide the same as ENZM or a polypeptide with at least one functional characteristic of ENZM. Included within this definition are polymorphisms which may or may not be readily detectable using a particular oligonucleotide probe of the polynucleotide encoding ENZM, and improper or unexpected hybridization to allelic variants,

with a locus other than the normal chromosomal locus for the polynucleotide encoding ENZM. The encoded protein may also be "altered," and may contain deletions, insertions, or substitutions of amino acid residues which produce a silent change and result in a functionally equivalent ENZM. Deliberate amino acid substitutions may be made on the basis of one or more similarities in polarity, charge, solubility, hydrophobicity, hydrophilicity, and/or the amphipathic nature of the residues, as long as the biological or immunological activity of ENZM is retained. For example, negatively charged amino acids may include aspartic acid and glutamic acid, and positively charged amino acids may include lysine and arginine. Amino acids with uncharged polar side chains having similar hydrophilicity values may include: asparagine and glutamine; and serine and threonine. Amino acids with uncharged side chains having similar hydrophilicity values may include: leucine, isoleucine, and valine; glycine and alanine; and phenylalanine and tyrosine.

The terms "amino acid" and "amino acid sequence" can refer to an oligopeptide, a peptide, a polypeptide, or a protein sequence, or a fragment of any of these, and to naturally occurring or synthetic molecules. Where "amino acid sequence" is recited to refer to a sequence of a naturally occurring protein molecule, "amino acid sequence" and like terms are not meant to limit the amino acid sequence to the complete native amino acid sequence associated with the recited protein molecule.

"Amplification" relates to the production of additional copies of a nucleic acid. Amplification may be carried out using polymerase chain reaction (PCR) technologies or other nucleic acid amplification technologies well known in the art.

The term "antagonist" refers to a molecule which inhibits or attenuates the biological activity of ENZM. Antagonists may include proteins such as antibodies, anticalins, nucleic acids, carbohydrates, small molecules, or any other compound or composition which modulates the activity of ENZM either by directly interacting with ENZM or by acting on components of the biological pathway in which ENZM participates.

The term "antibody" refers to intact immunoglobulin molecules as well as to fragments thereof, such as Fab, $F(ab')_2$, and Fv fragments, which are capable of binding an epitopic determinant. Antibodies that bind ENZM polypeptides can be prepared using intact polypeptides or using fragments containing small peptides of interest as the immunizing antigen. The polypeptide or oligopeptide used to immunize an animal (e.g., a mouse, a rat, or a rabbit) can be derived from the translation of RNA, or synthesized chemically, and can be conjugated to a carrier protein if desired. Commonly used carriers that are chemically coupled to peptides include bovine serum albumin, thyroglobulin, and keyhole limpet hemocyanin (KLH). The coupled peptide is then used to immunize the animal.

The term "antigenic determinant" refers to that region of a molecule (i.e., an epitope) that

makes contact with a particular antibody. When a protein or a fragment of a protein is used to immunize a host animal, numerous regions of the protein may induce the production of antibodies which bind specifically to antigenic determinants (particular regions or three-dimensional structures on the protein). An antigenic determinant may compete with the intact antigen (i.e., the immunogen used to elicit the immune response) for binding to an antibody.

The term "aptamer" refers to a nucleic acid or oligonucleotide molecule that binds to a specific molecular target. Aptamers are derived from an *in vitro* evolutionary process (e.g., SELEX (Systematic Evolution of Ligands by EXponential Enrichment), described in U.S. Patent No. 5,270,163), which selects for target-specific aptamer sequences from large combinatorial libraries. Aptamer compositions may be double-stranded or single-stranded, and may include deoxyribonucleotides, ribonucleotides, nucleotide derivatives, or other nucleotide-like molecules. The nucleotide components of an aptamer may have modified sugar groups (e.g., the 2'-OH group of a ribonucleotide may be replaced by 2'-F or 2'-NH_2), which may improve a desired property, e.g., resistance to nucleases or longer lifetime in blood. Aptamers may be conjugated to other molecules, e.g., a high molecular weight carrier to slow clearance of the aptamer from the circulatory system. Aptamers may be specifically cross-linked to their cognate ligands, e.g., by photo-activation of a cross-linker (Brody, E.N. and L. Gold (2000) J. Biotechnol. 74:5-13).

The term "intramer" refers to an aptamer which is expressed *in vivo*. For example, a vaccinia virus-based RNA expression system has been used to express specific RNA aptamers at high levels in the cytoplasm of leukocytes (Blind, M. et al. (1999) Proc. Natl. Acad. Sci. USA 96:3606-3610).

The term "spiegelmer" refers to an aptamer which includes L-DNA, L-RNA, or other left-handed nucleotide derivatives or nucleotide-like molecules. Aptamers containing left-handed nucleotides are resistant to degradation by naturally occurring enzymes, which normally act on substrates containing right-handed nucleotides.

The term "antisense" refers to any composition capable of base-pairing with the "sense" (coding) strand of a polynucleotide having a specific nucleic acid sequence. Antisense compositions may include DNA; RNA; peptide nucleic acid (PNA); oligonucleotides having modified backbone linkages such as phosphorothioates, methylphosphonates, or benzylphosphonates; oligonucleotides having modified sugar groups such as 2'-methoxyethyl sugars or 2'-methoxyethoxy sugars; or oligonucleotides having modified bases such as 5-methyl cytosine, 2'-deoxyuracil, or 7-deaza-2'-deoxyguanosine. Antisense molecules may be produced by any method including chemical synthesis or transcription. Once introduced into a cell, the complementary antisense molecule base-pairs with a naturally occurring nucleic acid sequence produced by the cell to form duplexes which block either transcription or translation. The designation "negative" or "minus" can refer to the antisense strand, and the designation "positive" or "plus" can refer to the sense strand of a reference DNA molecule.

The term "biologically active" refers to a protein having structural, regulatory, or biochemical functions of a naturally occurring molecule. Likewise, "immunologically active" or "immunogenic" refers to the capability of the natural, recombinant, or synthetic ENZM, or of any oligopeptide thereof, to induce a specific immune response in appropriate animals or cells and to bind with specific antibodies.

"Complementary" describes the relationship between two single-stranded nucleic acid sequences that anneal by base-pairing. For example, 5'-AGT-3' pairs with its complement, 3'-TCA-5'.

A "composition comprising a given polynucleotide" and a "composition comprising a given polypeptide" can refer to any composition containing the given polynucleotide or polypeptide. The composition may comprise a dry formulation or an aqueous solution. Compositions comprising polynucleotides encoding ENZM or fragments of ENZM may be employed as hybridization probes. The probes may be stored in freeze-dried form and may be associated with a stabilizing agent such as a carbohydrate. In hybridizations, the probe may be deployed in an aqueous solution containing salts (e.g., NaCl), detergents (e.g., sodium dodecyl sulfate; SDS), and other components (e.g., Denhardt's solution, dry milk, salmon sperm DNA, etc.).

"Consensus sequence" refers to a nucleic acid sequence which has been subjected to repeated DNA sequence analysis to resolve uncalled bases, extended using the XL-PCR kit (Applied Biosystems, Foster City CA) in the 5' and/or the 3' direction, and resequenced, or which has been assembled from one or more overlapping cDNA, EST, or genomic DNA fragments using a computer program for fragment assembly, such as the GELVIEW fragment assembly system (Accelrys, Burlington MA) or Phrap (University of Washington, Seattle WA). Some sequences have been both extended and assembled to produce the consensus sequence.

"Conservative amino acid substitutions" are those substitutions that are predicted to least interfere with the properties of the original protein, i.e., the structure and especially the function of the protein is conserved and not significantly changed by such substitutions. The table below shows amino acids which may be substituted for an original amino acid in a protein and which are regarded as conservative amino acid substitutions.

Original Residue	Conservative Substitution
Ala	Gly, Ser
Arg	His, Lys
Asn	Asp, Gln, His
Asp	Asn, Glu
Cys	Ala, Ser
Gln	Asn, Glu, His
Glu	Asp, Gln, His
Gly	Ala
His	Asn, Arg, Gln, Glu

Ile	Leu, Val
Leu	Ile, Val
Lys	Arg, Gln, Glu
Met	Leu, Ile
Phe	His, Met, Leu, Trp, Tyr
Ser	Cys, Thr
Thr	Ser, Val
Trp	Phe, Tyr
Tyr	His, Phe, Trp
Val	Ile, Leu, Thr

Conservative amino acid substitutions generally maintain (a) the structure of the polypeptide backbone in the area of the substitution, for example, as a beta sheet or alpha helical conformation, (b) the charge or hydrophobicity of the molecule at the site of the substitution, and/or (c) the bulk of the side chain.

A "deletion" refers to a change in the amino acid or nucleotide sequence that results in the absence of one or more amino acid residues or nucleotides.

The term "derivative" refers to a chemically modified polynucleotide or polypeptide. Chemical modifications of a polynucleotide can include, for example, replacement of hydrogen by an alkyl, acyl, hydroxyl, or amino group. A derivative polynucleotide encodes a polypeptide which retains at least one biological or immunological function of the natural molecule. A derivative polypeptide is one modified by glycosylation, pegylation, or any similar process that retains at least one biological or immunological function of the polypeptide from which it was derived.

A "detectable label" refers to a reporter molecule or enzyme that is capable of generating a measurable signal and is covalently or noncovalently joined to a polynucleotide or polypeptide.

"Differential expression" refers to increased or upregulated; or decreased, downregulated, or absent gene or protein expression, determined by comparing at least two different samples. Such comparisons may be carried out between, for example, a treated and an untreated sample, or a diseased and a normal sample.

"Exon shuffling" refers to the recombination of different coding regions (exons). Since an exon may represent a structural or functional domain of the encoded protein, new proteins may be assembled through the novel reassortment of stable substructures, thus allowing acceleration of the evolution of new protein functions.

A "fragment" is a unique portion of ENZM or a polynucleotide encoding ENZM which can be identical in sequence to, but shorter in length than, the parent sequence. A fragment may comprise up to the entire length of the defined sequence, minus one nucleotide/amino acid residue. For example, a fragment may comprise from about 5 to about 1000 contiguous nucleotides or amino acid residues. A fragment used as a probe, primer, antigen, therapeutic molecule, or for other purposes, may be at least 5, 10, 15, 16, 20, 25, 30, 40, 50, 60, 75, 100, 150, 250 or at least 500 contiguous

nucleotides or amino acid residues in length. Fragments may be preferentially selected from certain regions of a molecule. For example, a polypeptide fragment may comprise a certain length of contiguous amino acids selected from the first 250 or 500 amino acids (or first 25% or 50%) of a polypeptide as shown in a certain defined sequence. Clearly these lengths are exemplary, and any length that is supported by the specification, including the Sequence Listing, tables, and figures, may be encompassed by the present embodiments.

A fragment of SEQ ID NO:56-110 can comprise a region of unique polynucleotide sequence that specifically identifies SEQ ID NO:56-110, for example, as distinct from any other sequence in the genome from which the fragment was obtained. A fragment of SEQ ID NO:56-110 can be employed in one or more embodiments of methods of the invention, for example, in hybridization and amplification technologies and in analogous methods that distinguish SEQ ID NO:56-110 from related polynucleotides. The precise length of a fragment of SEQ ID NO:56-110 and the region of SEQ ID NO:56-110 to which the fragment corresponds are routinely determinable by one of ordinary skill in the art based on the intended purpose for the fragment.

A fragment of SEQ ID NO:1-55 is encoded by a fragment of SEQ ID NO:56-110. A fragment of SEQ ID NO:1-55 can comprise a region of unique amino acid sequence that specifically identifies SEQ ID NO:1-55. For example, a fragment of SEQ ID NO:1-55 can be used as an immunogenic peptide for the development of antibodies that specifically recognize SEQ ID NO:1-55. The precise length of a fragment of SEQ ID NO:1-55 and the region of SEQ ID NO:1-55 to which the fragment corresponds can be determined based on the intended purpose for the fragment using one or more analytical methods described herein or otherwise known in the art.

A "full length" polynucleotide is one containing at least a translation initiation codon (e.g., methionine) followed by an open reading frame and a translation termination codon. A "full length" polynucleotide sequence encodes a "full length" polypeptide sequence.

"Homology" refers to sequence similarity or, alternatively, sequence identity, between two or more polynucleotide sequences or two or more polypeptide sequences.

The terms "percent identity" and "% identity," as applied to polynucleotide sequences, refer to the percentage of identical nucleotide matches between at least two polynucleotide sequences aligned using a standardized algorithm. Such an algorithm may insert, in a standardized and reproducible way, gaps in the sequences being compared in order to optimize alignment between two sequences, and therefore achieve a more meaningful comparison of the two sequences.

Percent identity between polynucleotide sequences may be determined using one or more computer algorithms or programs known in the art or described herein. For example, percent identity can be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program. This program is part of the

LASERGENE software package, a suite of molecular biological analysis programs (DNASTAR, Madison WI). CLUSTAL V is described in Higgins, D.G. and P.M. Sharp (1989; CABIOS 5:151-153) and in Higgins, D.G. et al. (1992; CABIOS 8:189-191). For pairwise alignments of polynucleotide sequences, the default parameters are set as follows: Ktuple=2, gap penalty=5, window=4, and "diagonals saved"=4. The "weighted" residue weight table is selected as the default.

Alternatively, a suite of commonly used and freely available sequence comparison algorithms which can be used is provided by the National Center for Biotechnology Information (NCBI) Basic Local Alignment Search Tool (BLAST) (Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410), which is available from several sources, including the NCBI, Bethesda, MD, and on the Internet at http://www.ncbi.nlm.nih.gov/BLAST/. The BLAST software suite includes various sequence analysis programs including "blastn," that is used to align a known polynucleotide sequence with other polynucleotide sequences from a variety of databases. Also available is a tool called "BLAST 2 Sequences" that is used for direct pairwise comparison of two nucleotide sequences. "BLAST 2 Sequences" can be accessed and used interactively at http://www.ncbi.nlm.nih.gov/gorf/bl2.html. The "BLAST 2 Sequences" tool can be used for both blastn and blastp (discussed below). BLAST programs are commonly used with gap and other parameters set to default settings. For example, to compare two nucleotide sequences, one may use blastn with the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Reward for match: 1

Penalty for mismatch: -2

Open Gap: 5 and Extension Gap: 2 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 11

Filter: on

Percent identity may be measured over the length of an entire defined sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined sequence, for instance, a fragment of at least 20, at least 30, at least 40, at least 50, at least 70, at least 100, or at least 200 contiguous nucleotides. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures, or Sequence Listing, may be used to describe a length over which percentage identity may be measured.

Nucleic acid sequences that do not show a high degree of identity may nevertheless encode similar amino acid sequences due to the degeneracy of the genetic code. It is understood that changes

in a nucleic acid sequence can be made using this degeneracy to produce multiple nucleic acid sequences that all encode substantially the same protein.

The phrases "percent identity" and "% identity," as applied to polypeptide sequences, refer to the percentage of identical residue matches between at least two polypeptide sequences aligned using a standardized algorithm. Methods of polypeptide sequence alignment are well-known. Some alignment methods take into account conservative amino acid substitutions. Such conservative substitutions, explained in more detail above, generally preserve the charge and hydrophobicity at the site of substitution, thus preserving the structure (and therefore function) of the polypeptide. The phrases "percent similarity" and "% similarity," as applied to polypeptide sequences, refer to the percentage of residue matches, including identical residue matches and conservative substitutions, between at least two polypeptide sequences aligned using a standardized algorithm. In contrast, conservative substitutions are not included in the calculation of percent identity between polypeptide sequences.

Percent identity between polypeptide sequences may be determined using the default parameters of the CLUSTAL V algorithm as incorporated into the MEGALIGN version 3.12e sequence alignment program (described and referenced above). For pairwise alignments of polypeptide sequences using CLUSTAL V, the default parameters are set as follows: Ktuple=1, gap penalty=3, window=5, and "diagonals saved"=5. The PAM250 matrix is selected as the default residue weight table.

Alternatively the NCBI BLAST software suite may be used. For example, for a pairwise comparison of two polypeptide sequences, one may use the "BLAST 2 Sequences" tool Version 2.0.12 (April-21-2000) with blastp set at default parameters. Such default parameters may be, for example:

Matrix: BLOSUM62

Open Gap: 11 and Extension Gap: 1 penalties

Gap x drop-off: 50

Expect: 10

Word Size: 3

Filter: on

Percent identity may be measured over the length of an entire defined polypeptide sequence, for example, as defined by a particular SEQ ID number, or may be measured over a shorter length, for example, over the length of a fragment taken from a larger, defined polypeptide sequence, for instance, a fragment of at least 15, at least 20, at least 30, at least 40, at least 50, at least 70 or at least 150 contiguous residues. Such lengths are exemplary only, and it is understood that any fragment length supported by the sequences shown herein, in the tables, figures or Sequence Listing, may be

used to describe a length over which percentage identity may be measured.

"Human artificial chromosomes" (HACs) are linear microchromosomes which may contain DNA sequences of about 6 kb to 10 Mb in size and which contain all of the elements required for chromosome replication, segregation and maintenance.

The term "humanized antibody" refers to an antibody molecule in which the amino acid sequence in the non-antigen binding regions has been altered so that the antibody more closely resembles a human antibody, and still retains its original binding ability.

"Hybridization" refers to the process by which a polynucleotide strand anneals with a complementary strand through base pairing under defined hybridization conditions. Specific hybridization is an indication that two nucleic acid sequences share a high degree of complementarity. Specific hybridization complexes form under permissive annealing conditions and remain hybridized after the "washing" step(s). The washing step(s) is particularly important in determining the stringency of the hybridization process, with more stringent conditions allowing less non-specific binding, i.e., binding between pairs of nucleic acid strands that are not perfectly matched. Permissive conditions for annealing of nucleic acid sequences are routinely determinable by one of ordinary skill in the art and may be consistent among hybridization experiments, whereas wash conditions may be varied among experiments to achieve the desired stringency, and therefore hybridization specificity. Permissive annealing conditions occur, for example, at 68°C in the presence of about 6 x SSC, about 1% (w/v) SDS, and about 100 μg/ml sheared, denatured salmon sperm DNA.

Generally, stringency of hybridization is expressed, in part, with reference to the temperature under which the wash step is carried out. Such wash temperatures are typically selected to be about 5°C to 20°C lower than the thermal melting point (T_m) for the specific sequence at a defined ionic strength and pH. The T_m is the temperature (under defined ionic strength and pH) at which 50% of the target sequence hybridizes to a perfectly matched probe. An equation for calculating T_m and conditions for nucleic acid hybridization are well known and can be found in Sambrook, J. and D.W. Russell (2001; Molecular Cloning: A Laboratory Manual, 3rd ed., vol. 1-3, Cold Spring Harbor Press, Cold Spring Harbor NY, ch. 9).

High stringency conditions for hybridization between polynucleotides of the present invention include wash conditions of 68°C in the presence of about 0.2 x SSC and about 0.1% SDS, for 1 hour. Alternatively, temperatures of about 65°C, 60°C, 55°C, or 42°C may be used. SSC concentration may be varied from about 0.1 to 2 x SSC, with SDS being present at about 0.1%. Typically, blocking reagents are used to block non-specific hybridization. Such blocking reagents include, for instance, sheared and denatured salmon sperm DNA at about 100-200 μg/ml. Organic solvent, such as formamide at a concentration of about 35-50% v/v, may also be used under particular

circumstances, such as for RNA:DNA hybridizations. Useful variations on these wash conditions will be readily apparent to those of ordinary skill in the art. Hybridization, particularly under high stringency conditions, may be suggestive of evolutionary similarity between the nucleotides. Such similarity is strongly indicative of a similar role for the nucleotides and their encoded polypeptides.

The term "hybridization complex" refers to a complex formed between two nucleic acids by virtue of the formation of hydrogen bonds between complementary bases. A hybridization complex may be formed in solution (e.g., C_0t or R_0t analysis) or formed between one nucleic acid present in solution and another nucleic acid immobilized on a solid support (e.g., paper, membranes, filters, chips, pins or glass slides, or any other appropriate substrate to which cells or their nucleic acids have been fixed).

The words "insertion" and "addition" refer to changes in an amino acid or polynucleotide sequence resulting in the addition of one or more amino acid residues or nucleotides, respectively.

"Immune response" can refer to conditions associated with inflammation, trauma, immune disorders, or infectious or genetic disease, etc. These conditions can be characterized by expression of various factors, e.g., cytokines, chemokines, and other signaling molecules, which may affect cellular and systemic defense systems.

An "immunogenic fragment" is a polypeptide or oligopeptide fragment of ENZM which is capable of eliciting an immune response when introduced into a living organism, for example, a mammal. The term "immunogenic fragment" also includes any polypeptide or oligopeptide fragment of ENZM which is useful in any of the antibody production methods disclosed herein or known in the art.

The term "microarray" refers to an arrangement of a plurality of polynucleotides, polypeptides, antibodies, or other chemical compounds on a substrate.

The terms "element" and "array element" refer to a polynucleotide, polypeptide, antibody, or other chemical compound having a unique and defined position on a microarray.

The term "modulate" refers to a change in the activity of ENZM. For example, modulation may cause an increase or a decrease in protein activity, binding characteristics, or any other biological, functional, or immunological properties of ENZM.

The phrases "nucleic acid" and "nucleic acid sequence" refer to a nucleotide, oligonucleotide, polynucleotide, or any fragment thereof. These phrases also refer to DNA or RNA of genomic or synthetic origin which may be single-stranded or double-stranded and may represent the sense or the antisense strand, to peptide nucleic acid (PNA), or to any DNA-like or RNA-like material.

"Operably linked" refers to the situation in which a first nucleic acid sequence is placed in a functional relationship with a second nucleic acid sequence. For instance, a promoter is operably linked to a coding sequence if the promoter affects the transcription or expression of the coding

sequence. Operably linked DNA sequences may be in close proximity or contiguous and, where necessary to join two protein coding regions, in the same reading frame.

"Peptide nucleic acid" (PNA) refers to an antisense molecule or anti-gene agent which comprises an oligonucleotide of at least about 5 nucleotides in length linked to a peptide backbone of amino acid residues ending in lysine. The terminal lysine confers solubility to the composition. PNAs preferentially bind complementary single stranded DNA or RNA and stop transcript elongation, and may be pegylated to extend their lifespan in the cell.

"Post-translational modification" of an ENZM may involve lipidation, glycosylation, phosphorylation, acetylation, racemization, proteolytic cleavage, and other modifications known in the art. These processes may occur synthetically or biochemically. Biochemical modifications will vary by cell type depending on the enzymatic milieu of ENZM.

"Probe" refers to nucleic acids encoding ENZM, their complements, or fragments thereof, which are used to detect identical, allelic or related nucleic acids. Probes are isolated oligonucleotides or polynucleotides attached to a detectable label or reporter molecule. Typical labels include radioactive isotopes, ligands, chemiluminescent agents, and enzymes. "Primers" are short nucleic acids, usually DNA oligonucleotides, which may be annealed to a target polynucleotide by complementary base-pairing. The primer may then be extended along the target DNA strand by a DNA polymerase enzyme. Primer pairs can be used for amplification (and identification) of a nucleic acid, e.g., by the polymerase chain reaction (PCR).

Probes and primers as used in the present invention typically comprise at least 15 contiguous nucleotides of a known sequence. In order to enhance specificity, longer probes and primers may also be employed, such as probes and primers that comprise at least 20, 25, 30, 40, 50, 60, 70, 80, 90, 100, or at least 150 consecutive nucleotides of the disclosed nucleic acid sequences. Probes and primers may be considerably longer than these examples, and it is understood that any length supported by the specification, including the tables, figures, and Sequence Listing, may be used.

Methods for preparing and using probes and primers are described in, for example, Sambrook, J. and D.W. Russell (2001; Molecular Cloning: A Laboratory Manual, 3rd ed., vol. 1-3, Cold Spring Harbor Press, Cold Spring Harbor NY), Ausubel, F.M. et al. (1999; Short Protocols in Molecular Biology, 4th ed., John Wiley & Sons, New York NY), and Innis, M. et al. (1990; PCR Protocols, A Guide to Methods and Applications, Academic Press, San Diego CA). PCR primer pairs can be derived from a known sequence, for example, by using computer programs intended for that purpose such as Primer (Version 0.5, 1991, Whitehead Institute for Biomedical Research, Cambridge MA).

Oligonucleotides for use as primers are selected using software known in the art for such purpose. For example, OLIGO 4.06 software is useful for the selection of PCR primer pairs of up to

100 nucleotides each, and for the analysis of oligonucleotides and larger polynucleotides of up to 5,000 nucleotides from an input polynucleotide sequence of up to 32 kilobases. Similar primer selection programs have incorporated additional features for expanded capabilities. For example, the PrimOU primer selection program (available to the public from the Genome Center at University of Texas South West Medical Center, Dallas TX) is capable of choosing specific primers from megabase sequences and is thus useful for designing primers on a genome-wide scope. The Primer3 primer selection program (available to the public from the Whitehead Institute/MIT Center for Genome Research, Cambridge MA) allows the user to input a "mispriming library," in which sequences to avoid as primer binding sites are user-specified. Primer3 is useful, in particular, for the selection of oligonucleotides for microarrays. (The source code for the latter two primer selection programs may also be obtained from their respective sources and modified to meet the user's specific needs.) The PrimeGen program (available to the public from the UK Human Genome Mapping Project Resource Centre, Cambridge UK) designs primers based on multiple sequence alignments, thereby allowing selection of primers that hybridize to either the most conserved or least conserved regions of aligned nucleic acid sequences. Hence, this program is useful for identification of both unique and conserved oligonucleotides and polynucleotide fragments. The oligonucleotides and polynucleotide fragments identified by any of the above selection methods are useful in hybridization technologies, for example, as PCR or sequencing primers, microarray elements, or specific probes to identify fully or partially complementary polynucleotides in a sample of nucleic acids. Methods of oligonucleotide selection are not limited to those described above.

A "recombinant nucleic acid" is a nucleic acid that is not naturally occurring or has a sequence that is made by an artificial combination of two or more otherwise separated segments of sequence. This artificial combination is often accomplished by chemical synthesis or, more commonly, by the artificial manipulation of isolated segments of nucleic acids, e.g., by genetic engineering techniques such as those described in Sambrook and Russell (*supra*). The term recombinant includes nucleic acids that have been altered solely by addition, substitution, or deletion of a portion of the nucleic acid. Frequently, a recombinant nucleic acid may include a nucleic acid sequence operably linked to a promoter sequence. Such a recombinant nucleic acid may be part of a vector that is used, for example, to transform a cell.

Alternatively, such recombinant nucleic acids may be part of a viral vector, e.g., based on a vaccinia virus, that could be use to vaccinate a mammal wherein the recombinant nucleic acid is expressed, inducing a protective immunological response in the mammal.

A "regulatory element" refers to a nucleic acid sequence usually derived from untranslated regions of a gene and includes enhancers, promoters, introns, and 5' and 3' untranslated regions (UTRs). Regulatory elements interact with host or viral proteins which control transcription,

translation, or RNA stability.

"Reporter molecules" are chemical or biochemical moieties used for labeling a nucleic acid, amino acid, or antibody. Reporter molecules include radionuclides; enzymes; fluorescent, chemiluminescent, or chromogenic agents; substrates; cofactors; inhibitors; magnetic particles; and other moieties known in the art.

An "RNA equivalent," in reference to a DNA molecule, is composed of the same linear sequence of nucleotides as the reference DNA molecule with the exception that all occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The term "sample" is used in its broadest sense. A sample suspected of containing ENZM, nucleic acids encoding ENZM, or fragments thereof may comprise a bodily fluid; an extract from a cell, chromosome, organelle, or membrane isolated from a cell; a cell; genomic DNA, RNA, or cDNA, in solution or bound to a substrate; a tissue; a tissue print; etc.

The terms "specific binding" and "specifically binding" refer to that interaction between a protein or peptide and an agonist, an antibody, an antagonist, a small molecule, or any natural or synthetic binding composition. The interaction is dependent upon the presence of a particular structure of the protein, e.g., the antigenic determinant or epitope, recognized by the binding molecule. For example, if an antibody is specific for epitope "A," the presence of a polypeptide comprising the epitope A, or the presence of free unlabeled A, in a reaction containing free labeled A and the antibody will reduce the amount of labeled A that binds to the antibody.

The term "substantially purified" refers to nucleic acid or amino acid sequences that are removed from their natural environment and are isolated or separated, and are at least about 60% free, preferably at least about 75% free, and most preferably at least about 90% free from other components with which they are naturally associated.

A "substitution" refers to the replacement of one or more amino acid residues or nucleotides by different amino acid residues or nucleotides, respectively.

"Substrate" refers to any suitable rigid or semi-rigid support including membranes, filters, chips, slides, wafers, fibers, magnetic or nonmagnetic beads, gels, tubing, plates, polymers, microparticles and capillaries. The substrate can have a variety of surface forms, such as wells, trenches, pins, channels and pores, to which polynucleotides or polypeptides are bound.

A "transcript image" or "expression profile" refers to the collective pattern of gene expression by a particular cell type or tissue under given conditions at a given time.

"Transformation" describes a process by which exogenous DNA is introduced into a recipient cell. Transformation may occur under natural or artificial conditions according to various methods well known in the art, and may rely on any known method for the insertion of foreign nucleic acid

sequences into a prokaryotic or eukaryotic host cell. The method for transformation is selected based on the type of host cell being transformed and may include, but is not limited to, bacteriophage or viral infection, electroporation, heat shock, lipofection, and particle bombardment. The term "transformed cells" includes stably transformed cells in which the inserted DNA is capable of replication either as an autonomously replicating plasmid or as part of the host chromosome, as well as transiently transformed cells which express the inserted DNA or RNA for limited periods of time.

A "transgenic organism," as used herein, is any organism, including but not limited to animals and plants, in which one or more of the cells of the organism contains heterologous nucleic acid introduced by way of human intervention, such as by transgenic techniques well known in the art. The nucleic acid is introduced into the cell, directly or indirectly by introduction into a precursor of the cell, by way of deliberate genetic manipulation, such as by microinjection or by infection with a recombinant virus. In another embodiment, the nucleic acid can be introduced by infection with a recombinant viral vector, such as a lentiviral vector (Lois, C. et al. (2002) Science 295:868-872). The term genetic manipulation does not include classical cross-breeding, or *in vitro* fertilization, but rather is directed to the introduction of a recombinant DNA molecule. The transgenic organisms contemplated in accordance with the present invention include bacteria, cyanobacteria, fungi, plants and animals. The isolated DNA of the present invention can be introduced into the host by methods known in the art, for example infection, transfection, transformation or transconjugation. Techniques for transferring the DNA of the present invention into such organisms are widely known and provided in references such as Sambrook and Russell (*supra*).

A "variant" of a particular nucleic acid sequence is defined as a nucleic acid sequence having at least 40% sequence identity to the particular nucleic acid sequence over a certain length of one of the nucleic acid sequences using blastn with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of nucleic acids may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity over a certain defined length. A variant may be described as, for example, an "allelic" (as defined above), "splice," "species," or "polymorphic" variant. A splice variant may have significant identity to a reference molecule, but will generally have a greater or lesser number of polynucleotides due to alternate splicing during mRNA processing. The corresponding polypeptide may possess additional functional domains or lack domains that are present in the reference molecule. Species variants are polynucleotides that vary from one species to another. The resulting polypeptides will generally have significant amino acid identity relative to each other. A polymorphic variant is a variation in the polynucleotide sequence of a particular gene between individuals of a given species. Polymorphic variants also may encompass "single nucleotide

polymorphisms" (SNPs) in which the polynucleotide sequence varies by one nucleotide base. The presence of SNPs may be indicative of, for example, a certain population, a disease state, or a propensity for a disease state.

A "variant" of a particular polypeptide sequence is defined as a polypeptide sequence having at least 40% sequence identity or sequence similarity to the particular polypeptide sequence over a certain length of one of the polypeptide sequences using blastp with the "BLAST 2 Sequences" tool Version 2.0.9 (May-07-1999) set at default parameters. Such a pair of polypeptides may show, for example, at least 50%, at least 60%, at least 70%, at least 80%, at least 85%, at least 90%, at least 91%, at least 92%, at least 93%, at least 94%, at least 95%, at least 96%, at least 97%, at least 98%, or at least 99% or greater sequence identity or sequence similarity over a certain defined length of one of the polypeptides.

THE INVENTION

Various embodiments of the invention include new human enzymes (ENZM), the polynucleotides encoding ENZM, and the use of these compositions for the diagnosis, treatment, or prevention of autoimmune/inflammatory disorders, infectious disorders, immune deficiencies, disorders of metabolism, reproductive disorders, neurological disorders, cardiovascular disorders, eye disorders, and cell proliferative disorders, including cancer.

Table 1 summarizes the nomenclature for the full length polynucleotide and polypeptide embodiments of the invention. Each polynucleotide and its corresponding polypeptide are correlated to a single Incyte project identification number (Incyte Project ID). Each polypeptide sequence is denoted by both a polypeptide sequence identification number (Polypeptide SEQ ID NO:) and an Incyte polypeptide sequence number (Incyte Polypeptide ID) as shown. Each polynucleotide sequence is denoted by both a polynucleotide sequence identification number (Polynucleotide SEQ ID NO:) and an Incyte polynucleotide consensus sequence number (Incyte Polynucleotide ID) as shown. Column 6 shows the Incyte ID numbers of physical, full length clones corresponding to the polypeptide and polynucleotide sequences of the invention. The full length clones encode polypeptides which have at least 95% sequence identity to the polypeptide sequences shown in column 3.

Table 2 shows sequences with homology to polypeptide embodiments of the invention as identified by BLAST analysis against the GenBank protein (genpept) database and the PROTEOME database. Columns 1 and 2 show the polypeptide sequence identification number (Polypeptide SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for polypeptides of the invention. Column 3 shows the GenBank identification number (GenBank ID NO:) of the nearest GenBank homolog and the PROTEOME database identification numbers

(PROTEOME ID NO:) of the nearest PROTEOME database homologs. Column 4 shows the probability scores for the matches between each polypeptide and its homolog(s). Column 5 shows the annotation of the GenBank and PROTEOME database homolog(s) along with relevant citations where applicable, all of which are expressly incorporated by reference herein.

Table 3 shows various structural features of the polypeptides of the invention. Columns 1 and 2 show the polypeptide sequence identification number (SEQ ID NO:) and the corresponding Incyte polypeptide sequence number (Incyte Polypeptide ID) for each polypeptide of the invention. Column 3 shows the number of amino acid residues in each polypeptide. Column 4 shows potential phosphorylation sites, and column 5 shows potential glycosylation sites, as determined by the MOTIFS program of the GCG sequence analysis software package (Accelrys, Burlington MA). Column 6 shows amino acid residues comprising signature sequences, domains, and motifs. Column 7 shows analytical methods for protein structure/function analysis and in some cases, searchable databases to which the analytical methods were applied.

Together, Tables 2 and 3 summarize the properties of polypeptides of the invention, and these properties establish that the claimed polypeptides are enzymes. For example, SEQ ID NO:42 is 99% identical, from residue R46 to residue L473, and 95% identical, from residue M5 to residue V51, to human flavin-containing monooxygenase (GenBank ID g182671) as determined by the Basic Local Alignment Search Tool (BLAST). The BLAST probability score is 5.8e-255, which indicates the probability of obtaining the observed polypeptide sequence alignment by chance. SEQ ID NO:42 also has homology to proteins that are flavin-containing monooxygenases, enzymes which catalyze monooxygenation of xenobiotic soft nucleophiles, and are flavin-containing monooxygenases, as determined by BLAST analysis using the PROTEOME database. SEQ ID NO:42 also contains a flavin-binding monooxygenase-like domain as determined by searching for statistically significant matches in the hidden Markov model (HMM)-based PFAM database of conserved protein families/domains. (See Table 3.) Data from BLIMPS and PROFILESCAN analyses provide further corroborative evidence that SEQ ID NO:42 is a flavin-binding monooxygenase. SEQ ID NO:1-41 and SEQ ID NO:43-55 were analyzed and annotated in a similar manner. The algorithms and parameters for the analysis of SEQ ID NO:1-55 are described in Table 7.

As shown in Table 4, the full length polynucleotide embodiments were assembled using cDNA sequences or coding (exon) sequences derived from genomic DNA, or any combination of these two types of sequences. Column 1 lists the polynucleotide sequence identification number (Polynucleotide SEQ ID NO:), the corresponding Incyte polynucleotide consensus sequence number (Incyte ID) for each polynucleotide of the invention, and the length of each polynucleotide sequence in basepairs. Column 2 shows the nucleotide start (5′) and stop (3′) positions of the cDNA and/or genomic sequences used to assemble the full length polynucleotide embodiments, and of fragments of

the polynucleotides which are useful, for example, in hybridization or amplification technologies that identify SEQ ID NO:56-110 or that distinguish between SEQ ID NO:56-110 and related polynucleotides.

The polynucleotide fragments described in Column 2 of Table 4 may refer specifically, for example, to Incyte cDNAs derived from tissue-specific cDNA libraries or from pooled cDNA libraries. Alternatively, the polynucleotide fragments described in column 2 may refer to GenBank cDNAs or ESTs which contributed to the assembly of the full length polynucleotides. In addition, the polynucleotide fragments described in column 2 may identify sequences derived from the ENSEMBL (The Sanger Centre, Cambridge, UK) database (*i.e.*, those sequences including the designation "ENST"). Alternatively, the polynucleotide fragments described in column 2 may be derived from the NCBI RefSeq Nucleotide Sequence Records Database (*i.e.*, those sequences including the designation "NM" or "NT") or the NCBI RefSeq Protein Sequence Records (*i.e.*, those sequences including the designation "NP"). Alternatively, the polynucleotide fragments described in column 2 may refer to assemblages of both cDNA and Genscan-predicted exons brought together by an "exon stitching" algorithm. For example, a polynucleotide sequence identified as FL_*XXXXXX*_N_1_N_2_*YYYYY*_N_3_N_4 represents a "stitched" sequence in which *XXXXXX* is the identification number of the cluster of sequences to which the algorithm was applied, and *YYYYY* is the number of the prediction generated by the algorithm, and $N_{1,2,3...}$, if present, represent specific exons that may have been manually edited during analysis (See Example V). Alternatively, the polynucleotide fragments in column 2 may refer to assemblages of exons brought together by an "exon-stretching" algorithm. For example, a polynucleotide sequence identified as FL*XXXXXX*_g*AAAAA*_g*BBBBB*_1_*N* is a "stretched" sequence, with *XXXXXX* being the Incyte project identification number, g*AAAAA* being the GenBank identification number of the human genomic sequence to which the "exon-stretching" algorithm was applied, g*BBBBB* being the GenBank identification number or NCBI RefSeq identification number of the nearest GenBank protein homolog, and *N* referring to specific exons (See Example V). In instances where a RefSeq sequence was used as a protein homolog for the "exon-stretching" algorithm, a RefSeq identifier (denoted by "NM," "NP," or "NT") may be used in place of the GenBank identifier (*i.e.*, g*BBBBB*).

Alternatively, a prefix identifies component sequences that were hand-edited, predicted from genomic DNA sequences, or derived from a combination of sequence analysis methods. The following Table lists examples of component sequence prefixes and corresponding sequence analysis methods associated with the prefixes (see Example IV and Example V).

Prefix	Type of analysis and/or examples of programs

GNN, GFG, ENST	Exon prediction from genomic sequences using, for example, GENSCAN (Stanford University, CA, USA) or FGENES (Computer Genomics Group, The Sanger Centre, Cambridge, UK).
GBI	Hand-edited analysis of genomic sequences.
FL	Stitched or stretched genomic sequences (see Example V).
INCY	Full length transcript and exon prediction from mapping of EST sequences to the genome. Genomic location and EST composition data are combined to predict the exons and resulting transcript.

In some cases, Incyte cDNA coverage redundant with the sequence coverage shown in Table 4 was obtained to confirm the final consensus polynucleotide sequence, but the relevant Incyte cDNA identification numbers are not shown.

Table 5 shows the representative cDNA libraries for those full length polynucleotides which were assembled using Incyte cDNA sequences. The representative cDNA library is the Incyte cDNA library which is most frequently represented by the Incyte cDNA sequences which were used to assemble and confirm the above polynucleotides. The tissues and vectors which were used to construct the cDNA libraries shown in Table 5 are described in Table 6.

Table 8 shows single nucleotide polymorphisms (SNPs) found in polynucleotide sequences of the invention, along with allele frequencies in different human populations. Columns 1 and 2 show the polynucleotide sequence identification number (SEQ ID NO:) and the corresponding Incyte project identification number (PID) for polynucleotides of the invention. Column 3 shows the Incyte identification number for the EST in which the SNP was detected (EST ID), and column 4 shows the identification number for the SNP (SNP ID). Column 5 shows the position within the EST sequence at which the SNP is located (EST SNP), and column 6 shows the position of the SNP within the full-length polynucleotide sequence (CB1 SNP). Column 7 shows the allele found in the EST sequence. Columns 8 and 9 show the two alleles found at the SNP site. Column 10 shows the amino acid encoded by the codon including the SNP site, based upon the allele found in the EST. Columns 11-14 show the frequency of allele 1 in four different human populations. An entry of n/d (not detected) indicates that the frequency of allele 1 in the population was too low to be detected, while n/a (not available) indicates that the allele frequency was not determined for the population.

The invention also encompasses ENZM variants. Various embodiments of ENZM variants can have at least about 80%, at least about 90%, or at least about 95% amino acid sequence identity to the ENZM amino acid sequence, and can contain at least one functional or structural characteristic of ENZM.

Various embodiments also encompass polynucleotides which encode ENZM. In a particular

embodiment, the invention encompasses a polynucleotide sequence comprising a sequence selected from the group consisting of SEQ ID NO:56-110, which encodes ENZM. The polynucleotide sequences of SEQ ID NO:56-110, as presented in the Sequence Listing, embrace the equivalent RNA sequences, wherein occurrences of the nitrogenous base thymine are replaced with uracil, and the sugar backbone is composed of ribose instead of deoxyribose.

The invention also encompasses variants of a polynucleotide encoding ENZM. In particular, such a variant polynucleotide will have at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a polynucleotide encoding ENZM. A particular aspect of the invention encompasses a variant of a polynucleotide comprising a sequence selected from the group consisting of SEQ ID NO:56-110 which has at least about 70%, or alternatively at least about 85%, or even at least about 95% polynucleotide sequence identity to a nucleic acid sequence selected from the group consisting of SEQ ID NO:56-110. Any one of the polynucleotide variants described above can encode a polypeptide which contains at least one functional or structural characteristic of ENZM.

In addition, or in the alternative, a polynucleotide variant of the invention is a splice variant of a polynucleotide encoding ENZM. A splice variant may have portions which have significant sequence identity to a polynucleotide encoding ENZM, but will generally have a greater or lesser number of polynucleotides due to additions or deletions of blocks of sequence arising from alternate splicing during mRNA processing. A splice variant may have less than about 70%, or alternatively less than about 60%, or alternatively less than about 50% polynucleotide sequence identity to a polynucleotide encoding ENZM over its entire length; however, portions of the splice variant will have at least about 70%, or alternatively at least about 85%, or alternatively at least about 95%, or alternatively 100% polynucleotide sequence identity to portions of the polynucleotide encoding ENZM. For example, a polynucleotide comprising a sequence of SEQ ID NO:60 and a polynucleotide comprising a sequence of SEQ ID NO:61 are splice variants of each other; a polynucleotide comprising a sequence of SEQ ID NO:62 and a polynucleotide comprising a sequence of SEQ ID NO:63 are splice variants of each other; a polynucleotide comprising a sequence of SEQ ID NO:67 and a polynucleotide comprising a sequence of SEQ ID NO:101 are splice variants of each other; a polynucleotide comprising a sequence of SEQ ID NO:69 and a polynucleotide comprising a sequence of SEQ ID NO:105 are splice variants of each other; and a polynucleotide comprising a sequence of SEQ ID NO:73 and a polynucleotide comprising a sequence of SEQ ID NO:74 are splice variants of each other. Any one of the splice variants described above can encode a polypeptide which contains at least one functional or structural characteristic of ENZM.

It will be appreciated by those skilled in the art that as a result of the degeneracy of the genetic code, a multitude of polynucleotide sequences encoding ENZM, some bearing minimal

similarity to the polynucleotide sequences of any known and naturally occurring gene, may be produced. Thus, the invention contemplates each and every possible variation of polynucleotide sequence that could be made by selecting combinations based on possible codon choices. These combinations are made in accordance with the standard triplet genetic code as applied to the polynucleotide sequence of naturally occurring ENZM, and all such variations are to be considered as being specifically disclosed.

Although polynucleotides which encode ENZM and its variants are generally capable of hybridizing to polynucleotides encoding naturally occurring ENZM under appropriately selected conditions of stringency, it may be advantageous to produce polynucleotides encoding ENZM or its derivatives possessing a substantially different codon usage, e.g., inclusion of non-naturally occurring codons. Codons may be selected to increase the rate at which expression of the peptide occurs in a particular prokaryotic or eukaryotic host in accordance with the frequency with which particular codons are utilized by the host. Other reasons for substantially altering the nucleotide sequence encoding ENZM and its derivatives without altering the encoded amino acid sequences include the production of RNA transcripts having more desirable properties, such as a greater half-life, than transcripts produced from the naturally occurring sequence.

The invention also encompasses production of polynucleotides which encode ENZM and ENZM derivatives, or fragments thereof, entirely by synthetic chemistry. After production, the synthetic polynucleotide may be inserted into any of the many available expression vectors and cell systems using reagents well known in the art. Moreover, synthetic chemistry may be used to introduce mutations into a polynucleotide encoding ENZM or any fragment thereof.

Embodiments of the invention can also include polynucleotides that are capable of hybridizing to the claimed polynucleotides, and, in particular, to those having the sequences shown in SEQ ID NO:56-110 and fragments thereof, under various conditions of stringency (Wahl, G.M. and S.L. Berger (1987) Methods Enzymol. 152:399-407; Kimmel, A.R. (1987) Methods Enzymol. 152:507-511). Hybridization conditions, including annealing and wash conditions, are described in "Definitions."

Methods for DNA sequencing are well known in the art and may be used to practice any of the embodiments of the invention. The methods may employ such enzymes as the Klenow fragment of DNA polymerase I, SEQUENASE (US Biochemical, Cleveland OH), Taq polymerase (Applied Biosystems), thermostable T7 polymerase (Amersham Biosciences, Piscataway NJ), or combinations of polymerases and proofreading exonucleases such as those found in the ELONGASE amplification system (Invitrogen, Carlsbad CA). Preferably, sequence preparation is automated with machines such as the MICROLAB 2200 liquid transfer system (Hamilton, Reno NV), PTC200 thermal cycler (MJ Research, Watertown MA) and ABI CATALYST 800 thermal cycler (Applied Biosystems).

Sequencing is then carried out using either the ABI 373 or 377 DNA sequencing system (Applied Biosystems), the MEGABACE 1000 DNA sequencing system (Amersham Biosciences), or other systems known in the art. The resulting sequences are analyzed using a variety of algorithms which are well known in the art (Ausubel et al., *supra*, ch. 7; Meyers, R.A. (1995) Molecular Biology and Biotechnology, Wiley VCH, New York NY, pp. 856-853).

The nucleic acids encoding ENZM may be extended utilizing a partial nucleotide sequence and employing various PCR-based methods known in the art to detect upstream sequences, such as promoters and regulatory elements. For example, one method which may be employed, restriction-site PCR, uses universal and nested primers to amplify unknown sequence from genomic DNA within a cloning vector (Sarkar, G. (1993) PCR Methods Applic. 2:318-322). Another method, inverse PCR, uses primers that extend in divergent directions to amplify unknown sequence from a circularized template. The template is derived from restriction fragments comprising a known genomic locus and surrounding sequences (Triglia, T. et al. (1988) Nucleic Acids Res. 16:8186). A third method, capture PCR, involves PCR amplification of DNA fragments adjacent to known sequences in human and yeast artificial chromosome DNA (Lagerstrom, M. et al. (1991) PCR Methods Applic. 1:111-119). In this method, multiple restriction enzyme digestions and ligations may be used to insert an engineered double-stranded sequence into a region of unknown sequence before performing PCR. Other methods which may be used to retrieve unknown sequences are known in the art (Parker, J.D. et al. (1991) Nucleic Acids Res. 19:3055-3060). Additionally, one may use PCR, nested primers, and PROMOTERFINDER libraries (Clontech, Palo Alto CA) to walk genomic DNA. This procedure avoids the need to screen libraries and is useful in finding intron/exon junctions. For all PCR-based methods, primers may be designed using commercially available software, such as OLIGO 4.06 primer analysis software (National Biosciences, Plymouth MN) or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the template at temperatures of about 68°C to 72°C.

When screening for full length cDNAs, it is preferable to use libraries that have been size-selected to include larger cDNAs. In addition, random-primed libraries, which often include sequences containing the 5' regions of genes, are preferable for situations in which an oligo d(T) library does not yield a full-length cDNA. Genomic libraries may be useful for extension of sequence into 5' non-transcribed regulatory regions.

Capillary electrophoresis systems which are commercially available may be used to analyze the size or confirm the nucleotide sequence of sequencing or PCR products. In particular, capillary sequencing may employ flowable polymers for electrophoretic separation, four different nucleotide-specific, laser-stimulated fluorescent dyes, and a charge coupled device camera for detection of the emitted wavelengths. Output/light intensity may be converted to electrical signal using appropriate

software (e.g., GENOTYPER and SEQUENCE NAVIGATOR, Applied Biosystems), and the entire process from loading of samples to computer analysis and electronic data display may be computer controlled. Capillary electrophoresis is especially preferable for sequencing small DNA fragments which may be present in limited amounts in a particular sample.

In another embodiment of the invention, polynucleotides or fragments thereof which encode ENZM may be cloned in recombinant DNA molecules that direct expression of ENZM, or fragments or functional equivalents thereof, in appropriate host cells. Due to the inherent degeneracy of the genetic code, other polynucleotides which encode substantially the same or a functionally equivalent polypeptides may be produced and used to express ENZM.

The polynucleotides of the invention can be engineered using methods generally known in the art in order to alter ENZM-encoding sequences for a variety of purposes including, but not limited to, modification of the cloning, processing, and/or expression of the gene product. DNA shuffling by random fragmentation and PCR reassembly of gene fragments and synthetic oligonucleotides may be used to engineer the nucleotide sequences. For example, oligonucleotide-mediated site-directed mutagenesis may be used to introduce mutations that create new restriction sites, alter glycosylation patterns, change codon preference, produce splice variants, and so forth.

The nucleotides of the present invention may be subjected to DNA shuffling techniques such as MOLECULARBREEDING (Maxygen Inc., Santa Clara CA; described in U.S. Patent No. 5,837,458; Chang, C.-C. et al. (1999) Nat. Biotechnol. 17:793-797; Christians, F.C. et al. (1999) Nat. Biotechnol. 17:259-264; and Crameri, A. et al. (1996) Nat. Biotechnol. 14:315-319) to alter or improve the biological properties of ENZM, such as its biological or enzymatic activity or its ability to bind to other molecules or compounds. DNA shuffling is a process by which a library of gene variants is produced using PCR-mediated recombination of gene fragments. The library is then subjected to selection or screening procedures that identify those gene variants with the desired properties. These preferred variants may then be pooled and further subjected to recursive rounds of DNA shuffling and selection/screening. Thus, genetic diversity is created through "artificial" breeding and rapid molecular evolution. For example, fragments of a single gene containing random point mutations may be recombined, screened, and then reshuffled until the desired properties are optimized. Alternatively, fragments of a given gene may be recombined with fragments of homologous genes in the same gene family, either from the same or different species, thereby maximizing the genetic diversity of multiple naturally occurring genes in a directed and controllable manner.

In another embodiment, polynucleotides encoding ENZM may be synthesized, in whole or in part, using one or more chemical methods well known in the art (Caruthers, M.H. et al. (1980) Nucleic Acids Symp. Ser. 7:215-223; Horn, T. et al. (1980) Nucleic Acids Symp. Ser. 7:225-232).

Alternatively, ENZM itself or a fragment thereof may be synthesized using chemical methods known in the art. For example, peptide synthesis can be performed using various solution-phase or solid-phase techniques (Creighton, T. (1984) Proteins, Structures and Molecular Properties, WH Freeman, New York NY, pp. 55-60; Roberge, J.Y. et al. (1995) Science 269:202-204). Automated synthesis may be achieved using the ABI 431A peptide synthesizer (Applied Biosystems). Additionally, the amino acid sequence of ENZM, or any part thereof, may be altered during direct synthesis and/or combined with sequences from other proteins, or any part thereof, to produce a variant polypeptide or a polypeptide having a sequence of a naturally occurring polypeptide.

The peptide may be substantially purified by preparative high performance liquid chromatography (Chiez, R.M. and F.Z. Regnier (1990) Methods Enzymol. 182:392-421). The composition of the synthetic peptides may be confirmed by amino acid analysis or by sequencing (Creighton, *supra*, pp. 28-53).

In order to express a biologically active ENZM, the polynucleotides encoding ENZM or derivatives thereof may be inserted into an appropriate expression vector, i.e., a vector which contains the necessary elements for transcriptional and translational control of the inserted coding sequence in a suitable host. These elements include regulatory sequences, such as enhancers, constitutive and inducible promoters, and 5' and 3' untranslated regions in the vector and in polynucleotides encoding ENZM. Such elements may vary in their strength and specificity. Specific initiation signals may also be used to achieve more efficient translation of polynucleotides encoding ENZM. Such signals include the ATG initiation codon and adjacent sequences, e.g. the Kozak sequence. In cases where a polynucleotide sequence encoding ENZM and its initiation codon and upstream regulatory sequences are inserted into the appropriate expression vector, no additional transcriptional or translational control signals may be needed. However, in cases where only coding sequence, or a fragment thereof, is inserted, exogenous translational control signals including an in-frame ATG initiation codon should be provided by the vector. Exogenous translational elements and initiation codons may be of various origins, both natural and synthetic. The efficiency of expression may be enhanced by the inclusion of enhancers appropriate for the particular host cell system used (Scharf, D. et al. (1994) Results Probl. Cell Differ. 20:125-162).

Methods which are well known to those skilled in the art may be used to construct expression vectors containing polynucleotides encoding ENZM and appropriate transcriptional and translational control elements. These methods include *in vitro* recombinant DNA techniques, synthetic techniques, and *in vivo* genetic recombination (Sambrook and Russell, *supra*, ch. 1-4, and 8; Ausubel et al., *supra*, ch. 1, 3, and 15).

A variety of expression vector/host systems may be utilized to contain and express polynucleotides encoding ENZM. These include, but are not limited to, microorganisms such as

bacteria transformed with recombinant bacteriophage, plasmid, or cosmid DNA expression vectors; yeast transformed with yeast expression vectors; insect cell systems infected with viral expression vectors (e.g., baculovirus); plant cell systems transformed with viral expression vectors (e.g., cauliflower mosaic virus, CaMV, or tobacco mosaic virus, TMV) or with bacterial expression vectors (e.g., Ti or pBR322 plasmids); or animal cell systems (Sambrook and Russell, *supra*; Ausubel et al., *supra*; Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509; Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945; Takamatsu, N. (1987) EMBO J. 6:307-311; The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196; Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659; Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355). Expression vectors derived from retroviruses, adenoviruses, or herpes or vaccinia viruses, or from various bacterial plasmids, may be used for delivery of polynucleotides to the targeted organ, tissue, or cell population (Di Nicola, M. et al. (1998) Cancer Gen. Ther. 5:350-356; Yu, M. et al. (1993) Proc. Natl. Acad. Sci. USA 90:6340-6344; Buller, R.M. et al. (1985) Nature 317:813-815; McGregor, D.P. et al. (1994) Mol. Immunol. 31:219-226; Verma, I.M. and N. Somia (1997) Nature 389:239-242). The invention is not limited by the host cell employed.

In bacterial systems, a number of cloning and expression vectors may be selected depending upon the use intended for polynucleotides encoding ENZM. For example, routine cloning, subcloning, and propagation of polynucleotides encoding ENZM can be achieved using a multifunctional *E. coli* vector such as PBLUESCRIPT (Stratagene, La Jolla CA) or PSPORT1 plasmid (Invitrogen). Ligation of polynucleotides encoding ENZM into the vector's multiple cloning site disrupts the *lacZ* gene, allowing a colorimetric screening procedure for identification of transformed bacteria containing recombinant molecules. In addition, these vectors may be useful for *in vitro* transcription, dideoxy sequencing, single strand rescue with helper phage, and creation of nested deletions in the cloned sequence (Van Heeke, G. and S.M. Schuster (1989) J. Biol. Chem. 264:5503-5509). When large quantities of ENZM are needed, e.g. for the production of antibodies, vectors which direct high level expression of ENZM may be used. For example, vectors containing the strong, inducible SP6 or T7 bacteriophage promoter may be used.

Yeast expression systems may be used for production of ENZM. A number of vectors containing constitutive or inducible promoters, such as alpha factor, alcohol oxidase, and PGH promoters, may be used in the yeast *Saccharomyces cerevisiae* or *Pichia pastoris*. In addition, such vectors direct either the secretion or intracellular retention of expressed proteins and enable integration of foreign polynucleotide sequences into the host genome for stable propagation (Ausubel et al., *supra*; Bitter, G.A. et al. (1987) Methods Enzymol. 153:516-544; Scorer, C.A. et al. (1994) Bio/Technology 12:181-184).

Plant systems may also be used for expression of ENZM. Transcription of polynucleotides encoding ENZM may be driven by viral promoters, e.g., the 35S and 19S promoters of CaMV used alone or in combination with the omega leader sequence from TMV (Takamatsu, N. (1987) EMBO J. 6:307-311). Alternatively, plant promoters such as the small subunit of RUBISCO or heat shock promoters may be used (Coruzzi, G. et al. (1984) EMBO J. 3:1671-1680; Broglie, R. et al. (1984) Science 224:838-843; Winter, J. et al. (1991) Results Probl. Cell Differ. 17:85-105). These constructs can be introduced into plant cells by direct DNA transformation or pathogen-mediated transfection (The McGraw Hill Yearbook of Science and Technology (1992) McGraw Hill, New York NY, pp. 191-196).

In mammalian cells, a number of viral-based expression systems may be utilized. In cases where an adenovirus is used as an expression vector, polynucleotides encoding ENZM may be ligated into an adenovirus transcription/translation complex consisting of the late promoter and tripartite leader sequence. Insertion in a non-essential E1 or E3 region of the viral genome may be used to obtain infective virus which expresses ENZM in host cells (Logan, J. and T. Shenk (1984) Proc. Natl. Acad. Sci. USA 81:3655-3659). In addition, transcription enhancers, such as the Rous sarcoma virus (RSV) enhancer, may be used to increase expression in mammalian host cells. SV40 or EBV-based vectors may also be used for high-level protein expression.

Human artificial chromosomes (HACs) may also be employed to deliver larger fragments of DNA than can be contained in and expressed from a plasmid. HACs of about 6 kb to 10 Mb are constructed and delivered via conventional delivery methods (liposomes, polycationic amino polymers, or vesicles) for therapeutic purposes (Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355).

For long term production of recombinant proteins in mammalian systems, stable expression of ENZM in cell lines is preferred. For example, polynucleotides encoding ENZM can be transformed into cell lines using expression vectors which may contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate vector. Following the introduction of the vector, cells may be allowed to grow for about 1 to 2 days in enriched media before being switched to selective media. The purpose of the selectable marker is to confer resistance to a selective agent, and its presence allows growth and recovery of cells which successfully express the introduced sequences. Resistant clones of stably transformed cells may be propagated using tissue culture techniques appropriate to the cell type.

Any number of selection systems may be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase and adenine phosphoribosyltransferase genes, for use in *tk*$^-$ and *apr*$^-$ cells, respectively (Wigler, M. et al. (1977) Cell 11:223-232; Lowy, I. et al. (1980) Cell 22:817-823). Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to

methotrexate; *neo* confers resistance to the aminoglycosides neomycin and G-418; and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively (Wigler, M. et al. (1980) Proc. Natl. Acad. Sci. USA 77:3567-3570; Colbere-Garapin, F. et al. (1981) J. Mol. Biol. 150:1-14). Additional selectable genes have been described, e.g., *trpB* and *hisD*, which alter cellular requirements for metabolites (Hartman, S.C. and R.C. Mulligan (1988) Proc. Natl. Acad. Sci. USA 85:8047-8051). Visible markers, e.g., anthocyanins, green fluorescent proteins (GFP; Clontech), β-glucuronidase and its substrate β-glucuronide, or luciferase and its substrate luciferin may be used. These markers can be used not only to identify transformants, but also to quantify the amount of transient or stable protein expression attributable to a specific vector system (Rhodes, C.A. (1995) Methods Mol. Biol. 55:121-131).

Although the presence/absence of marker gene expression suggests that the gene of interest is also present, the presence and expression of the gene may need to be confirmed. For example, if the sequence encoding ENZM is inserted within a marker gene sequence, transformed cells containing polynucleotides encoding ENZM can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding ENZM under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of the tandem gene as well.

In general, host cells that contain the polynucleotide encoding ENZM and that express ENZM may be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations, PCR amplification, and protein bioassay or immunoassay techniques which include membrane, solution, or chip based technologies for the detection and/or quantification of nucleic acid or protein sequences.

Immunological methods for detecting and measuring the expression of ENZM using either specific polyclonal or monoclonal antibodies are known in the art. Examples of such techniques include enzyme-linked immunosorbent assays (ELISAs), radioimmunoassays (RIAs), and fluorescence activated cell sorting (FACS). A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering epitopes on ENZM is preferred, but a competitive binding assay may be employed. These and other assays are well known in the art (Hampton, R. et al. (1990) Serological Methods, a Laboratory Manual, APS Press, St. Paul MN, Sect. IV; Coligan, J.E. et al. (1997) Current Protocols in Immunology, Greene Pub. Associates and Wiley-Interscience, New York NY; Pound, J.D. (1998) Immunochemical Protocols, Humana Press, Totowa NJ).

A wide variety of labels and conjugation techniques are known by those skilled in the art and may be used in various nucleic acid and amino acid assays. Means for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides encoding ENZM

include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, polynucleotides encoding ENZM, or any fragments thereof, may be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by addition of an appropriate RNA polymerase such as T7, T3, or SP6 and labeled nucleotides. These procedures may be conducted using a variety of commercially available kits, such as those provided by Amersham Biosciences, Promega (Madison WI), and US Biochemical. Suitable reporter molecules or labels which may be used for ease of detection include radionuclides, enzymes, fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

Host cells transformed with polynucleotides encoding ENZM may be cultured under conditions suitable for the expression and recovery of the protein from cell culture. The protein produced by a transformed cell may be secreted or retained intracellularly depending on the sequence and/or the vector used. As will be understood by those of skill in the art, expression vectors containing polynucleotides which encode ENZM may be designed to contain signal sequences which direct secretion of ENZM through a prokaryotic or eukaryotic cell membrane.

In addition, a host cell strain may be chosen for its ability to modulate expression of the inserted polynucleotides or to process the expressed protein in the desired fashion. Such modifications of the polypeptide include, but are not limited to, acetylation, carboxylation, glycosylation, phosphorylation, lipidation, and acylation. Post-translational processing which cleaves a "prepro" or "pro" form of the protein may also be used to specify protein targeting, folding, and/or activity. Different host cells which have specific cellular machinery and characteristic mechanisms for post-translational activities (e.g., CHO, HeLa, MDCK, HEK293, and WI38) are available from the American Type Culture Collection (ATCC, Manassas VA) and may be chosen to ensure the correct modification and processing of the foreign protein.

In another embodiment of the invention, natural, modified, or recombinant polynucleotides encoding ENZM may be ligated to a heterologous sequence resulting in translation of a fusion protein in any of the aforementioned host systems. For example, a chimeric ENZM protein containing a heterologous moiety that can be recognized by a commercially available antibody may facilitate the screening of peptide libraries for inhibitors of ENZM activity. Heterologous protein and peptide moieties may also facilitate purification of fusion proteins using commercially available affinity matrices. Such moieties include, but are not limited to, glutathione S-transferase (GST), maltose binding protein (MBP), thioredoxin (Trx), calmodulin binding peptide (CBP), 6-His, FLAG, *c-myc*, and hemagglutinin (HA). GST, MBP, Trx, CBP, and 6-His enable purification of their cognate fusion proteins on immobilized glutathione, maltose, phenylarsine oxide, calmodulin, and metal-chelate resins, respectively. FLAG, *c-myc*, and hemagglutinin (HA) enable immunoaffinity purification of

fusion proteins using commercially available monoclonal and polyclonal antibodies that specifically recognize these epitope tags. A fusion protein may also be engineered to contain a proteolytic cleavage site located between the ENZM encoding sequence and the heterologous protein sequence, so that ENZM may be cleaved away from the heterologous moiety following purification. Methods for fusion protein expression and purification are discussed in Ausubel et al. (*supra*, ch. 10 and 16). A variety of commercially available kits may also be used to facilitate expression and purification of fusion proteins.

In another embodiment, synthesis of radiolabeled ENZM may be achieved *in vitro* using the TNT rabbit reticulocyte lysate or wheat germ extract system (Promega). These systems couple transcription and translation of protein-coding sequences operably associated with the T7, T3, or SP6 promoters. Translation takes place in the presence of a radiolabeled amino acid precursor, for example, ^{35}S-methionine.

ENZM, fragments of ENZM, or variants of ENZM may be used to screen for compounds that specifically bind to ENZM. One or more test compounds may be screened for specific binding to ENZM. In various embodiments, 1, 2, 3, 4, 5, 10, 20, 50, 100, or 200 test compounds can be screened for specific binding to ENZM. Examples of test compounds can include antibodies, anticalins, oligonucleotides, proteins (e.g., ligands or receptors), or small molecules.

In related embodiments, variants of ENZM can be used to screen for binding of test compounds, such as antibodies, to ENZM, a variant of ENZM, or a combination of ENZM and/or one or more variants ENZM. In an embodiment, a variant of ENZM can be used to screen for compounds that bind to a variant of ENZM, but not to ENZM having the exact sequence of a sequence of SEQ ID NO:1-55. ENZM variants used to perform such screening can have a range of about 50% to about 99% sequence identity to ENZM, with various embodiments having 60%, 70%, 75%, 80%, 85%, 90%, and 95% sequence identity.

In an embodiment, a compound identified in a screen for specific binding to ENZM can be closely related to the natural ligand of ENZM, e.g., a ligand or fragment thereof, a natural substrate, a structural or functional mimetic, or a natural binding partner (Coligan, J.E. et al. (1991) Current Protocols in Immunology 1(2):Chapter 5). In another embodiment, the compound thus identified can be a natural ligand of a receptor ENZM (Howard, A.D. et al. (2001) Trends Pharmacol. Sci.22:132-140; Wise, A. et al. (2002) Drug Discovery Today 7:235-246).

In other embodiments, a compound identified in a screen for specific binding to ENZM can be closely related to the natural receptor to which ENZM binds, at least a fragment of the receptor, or a fragment of the receptor including all or a portion of the ligand binding site or binding pocket. For example, the compound may be a receptor for ENZM which is capable of propagating a signal, or a decoy receptor for ENZM which is not capable of propagating a signal (Ashkenazi, A. and V.M.

Divit (1999) Curr. Opin. Cell Biol. 11:255-260; Mantovani, A. et al. (2001) Trends Immunol. 22:328-336). The compound can be rationally designed using known techniques. Examples of such techniques include those used to construct the compound etanercept (ENBREL; Amgen Inc., Thousand Oaks CA), which is efficacious for treating rheumatoid arthritis in humans. Etanercept is an engineered p75 tumor necrosis factor (TNF) receptor dimer linked to the Fc portion of human IgG_1 (Taylor, P.C. et al. (2001) Curr. Opin. Immunol. 13:611-616).

In one embodiment, two or more antibodies having similar or, alternatively, different specificities can be screened for specific binding to ENZM, fragments of ENZM, or variants of ENZM. The binding specificity of the antibodies thus screened can thereby be selected to identify particular fragments or variants of ENZM. In one embodiment, an antibody can be selected such that its binding specificity allows for preferential identification of specific fragments or variants of ENZM. In another embodiment, an antibody can be selected such that its binding specificity allows for preferential diagnosis of a specific disease or condition having increased, decreased, or otherwise abnormal production of ENZM.

In an embodiment, anticalins can be screened for specific binding to ENZM, fragments of ENZM, or variants of ENZM. Anticalins are ligand-binding proteins that have been constructed based on a lipocalin scaffold (Weiss, G.A. and H.B. Lowman (2000) Chem. Biol. 7:R177-R184; Skerra, A. (2001) J. Biotechnol. 74:257-275). The protein architecture of lipocalins can include a beta-barrel having eight antiparallel beta-strands, which supports four loops at its open end. These loops form the natural ligand-binding site of the lipocalins, a site which can be re-engineered *in vitro* by amino acid substitutions to impart novel binding specificities. The amino acid substitutions can be made using methods known in the art or described herein, and can include conservative substitutions (e.g., substitutions that do not alter binding specificity) or substitutions that modestly, moderately, or significantly alter binding specificity.

In one embodiment, screening for compounds which specifically bind to, stimulate, or inhibit ENZM involves producing appropriate cells which express ENZM, either as a secreted protein or on the cell membrane. Preferred cells can include cells from mammals, yeast, *Drosophila*, or *E. coli*. Cells expressing ENZM or cell membrane fractions which contain ENZM are then contacted with a test compound and binding, stimulation, or inhibition of activity of either ENZM or the compound is analyzed.

An assay may simply test binding of a test compound to the polypeptide, wherein binding is detected by a fluorophore, radioisotope, enzyme conjugate, or other detectable label. For example, the assay may comprise the steps of combining at least one test compound with ENZM, either in solution or affixed to a solid support, and detecting the binding of ENZM to the compound. Alternatively, the assay may detect or measure binding of a test compound in the presence of a

labeled competitor. Additionally, the assay may be carried out using cell-free preparations, chemical libraries, or natural product mixtures, and the test compound(s) may be free in solution or affixed to a solid support.

An assay can be used to assess the ability of a compound to bind to its natural ligand and/or to inhibit the binding of its natural ligand to its natural receptors. Examples of such assays include radio-labeling assays such as those described in U.S. Patent No. 5,914,236 and U.S. Patent No. 6,372,724. In a related embodiment, one or more amino acid substitutions can be introduced into a polypeptide compound (such as a receptor) to improve or alter its ability to bind to its natural ligands (Matthews, D.J. and J.A. Wells. (1994) Chem. Biol. 1:25-30). In another related embodiment, one or more amino acid substitutions can be introduced into a polypeptide compound (such as a ligand) to improve or alter its ability to bind to its natural receptors (Cunningham, B.C. and J.A. Wells (1991) Proc. Natl. Acad. Sci. USA 88:3407-3411; Lowman, H.B. et al. (1991) J. Biol. Chem. 266:10982-10988).

ENZM, fragments of ENZM, or variants of ENZM may be used to screen for compounds that modulate the activity of ENZM. Such compounds may include agonists, antagonists, or partial or inverse agonists. In one embodiment, an assay is performed under conditions permissive for ENZM activity, wherein ENZM is combined with at least one test compound, and the activity of ENZM in the presence of a test compound is compared with the activity of ENZM in the absence of the test compound. A change in the activity of ENZM in the presence of the test compound is indicative of a compound that modulates the activity of ENZM. Alternatively, a test compound is combined with an *in vitro* or cell-free system comprising ENZM under conditions suitable for ENZM activity, and the assay is performed. In either of these assays, a test compound which modulates the activity of ENZM may do so indirectly and need not come in direct contact with the test compound. At least one and up to a plurality of test compounds may be screened.

In another embodiment, polynucleotides encoding ENZM or their mammalian homologs may be "knocked out" in an animal model system using homologous recombination in embryonic stem (ES) cells. Such techniques are well known in the art and are useful for the generation of animal models of human disease (see, e.g., U.S. Patent No. 5,175,383 and U.S. Patent No. 5,767,337). For example, mouse ES cells, such as the mouse 129/SvJ cell line, are derived from the early mouse embryo and grown in culture. The ES cells are transformed with a vector containing the gene of interest disrupted by a marker gene, e.g., the neomycin phosphotransferase gene (*neo*; Capecchi, M.R. (1989) Science 244:1288-1292). The vector integrates into the corresponding region of the host genome by homologous recombination. Alternatively, homologous recombination takes place using the Cre-loxP system to knockout a gene of interest in a tissue- or developmental stage-specific manner (Marth, J.D. (1996) Clin. Invest. 97:1999-2002; Wagner, K.U. et al. (1997) Nucleic Acids

Res. 25:4323-4330). Transformed ES cells are identified and microinjected into mouse cell blastocysts such as those from the C57BL/6 mouse strain. The blastocysts are surgically transferred to pseudopregnant dams, and the resulting chimeric progeny are genotyped and bred to produce heterozygous or homozygous strains. Transgenic animals thus generated may be tested with potential therapeutic or toxic agents.

Polynucleotides encoding ENZM may also be manipulated *in vitro* in ES cells derived from human blastocysts. Human ES cells have the potential to differentiate into at least eight separate cell lineages including endoderm, mesoderm, and ectodermal cell types. These cell lineages differentiate into, for example, neural cells, hematopoietic lineages, and cardiomyocytes (Thomson, J.A. et al. (1998) Science 282:1145-1147).

Polynucleotides encoding ENZM can also be used to create "knockin" humanized animals (pigs) or transgenic animals (mice or rats) to model human disease. With knockin technology, a region of a polynucleotide encoding ENZM is injected into animal ES cells, and the injected sequence integrates into the animal cell genome. Transformed cells are injected into blastulae, and the blastulae are implanted as described above. Transgenic progeny or inbred lines are studied and treated with potential pharmaceutical agents to obtain information on treatment of a human disease. Alternatively, a mammal inbred to overexpress ENZM, e.g., by secreting ENZM in its milk, may also serve as a convenient source of that protein (Janne, J. et al. (1998) Biotechnol. Annu. Rev. 4:55-74).

THERAPEUTICS

Chemical and structural similarity, e.g., in the context of sequences and motifs, exists between regions of ENZM and enzymes. In addition, examples of tissues expressing ENZM can be found in Table 6 and can also be found in Example XI. Therefore, ENZM appears to play a role in autoimmune/inflammatory disorders, infectious disorders, immune deficiencies, disorders of metabolism, reproductive disorders, neurological disorders, cardiovascular disorders, eye disorders, and cell proliferative disorders, including cancer. In the treatment of disorders associated with increased ENZM expression or activity, it is desirable to decrease the expression or activity of ENZM. In the treatment of disorders associated with decreased ENZM expression or activity, it is desirable to increase the expression or activity of ENZM.

Therefore, in one embodiment, ENZM or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of ENZM. Examples of such disorders include, but are not limited to, an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact

dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis, psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, and trauma; an infectious disorder such as a viral infection, e.g., caused by an adenovirus (acute respiratory disease, pneumonia), an arenavirus (lymphocytic choriomeningitis), a bunyavirus (Hantavirus), a coronavirus (pneumonia, chronic bronchitis), a hepadnavirus (hepatitis), a herpesvirus (herpes simplex virus, varicella-zoster virus, Epstein-Barr virus, cytomegalovirus), a flavivirus (yellow fever), an orthomyxovirus (influenza), a papillomavirus (cancer), a paramyxovirus (measles, mumps), a picornovirus (rhinovirus, poliovirus, coxsackie-virus), a polyomavirus (BK virus, JC virus), a poxvirus (smallpox), a reovirus (Colorado tick fever), a retrovirus (human immunodeficiency virus, human T lymphotropic virus), a rhabdovirus (rabies), a rotavirus (gastroenteritis), and a togavirus (encephalitis, rubella), and a bacterial infection, a fungal infection, a parasitic infection, a protozoal infection, and a helminthic infection; an immune deficiency, such as acquired immunodeficiency syndrome (AIDS), X-linked agammaglobinemia of Bruton, common variable immunodeficiency (CVI), DiGeorge's syndrome (thymic hypoplasia), thymic dysplasia, isolated IgA deficiency, severe combined immunodeficiency disease (SCID), immunodeficiency with thrombocytopenia and eczema (Wiskott-Aldrich syndrome), Chediak-Higashi syndrome, chronic granulomatous diseases, hereditary angioneurotic edema, and immunodeficiency associated with Cushing's disease; a disorder of metabolism such as Addison's disease, cerebrotendinous xanthomatosis, congenital adrenal hyperplasia, coumarin resistance, cystic fibrosis, diabetes, fatty hepatocirrhosis, fructose-1,6-diphosphatase deficiency, galactosemia, goiter, glucagonoma, glycogen storage diseases, hereditary fructose intolerance, hyperadrenalism, hypoadrenalism, hyperparathyroidism, hypoparathyroidism, hypercholesterolemia, hyperthyroidism, hypoglycemia, hypothyroidism, hyperlipidemia, hyperlipemia, a lipid myopathy, a lipodystrophy, a lysosomal storage disease, mannosidosis, neuraminidase deficiency, obesity, pentosuria phenylketonuria, pseudovitamin D-deficiency rickets; a reproductive disorder such as a disorder of prolactin production, infertility, including tubal disease, ovulatory defects, and endometriosis, a disruption of the estrous cycle, a disruption of the menstrual cycle, polycystic ovary syndrome, ovarian hyperstimulation syndrome, endometrial and ovarian tumors, uterine fibroids, autoimmune disorders, ectopic pregnancies, and teratogenesis, cancer of the breast, fibrocystic breast disease, and

galactorrhea, disruptions of spermatogenesis, abnormal sperm physiology, cancer of the testis, cancer of the prostate, benign prostatic hyperplasia, prostatitis, Peyronie's disease, impotence, carcinoma of the male breast, and gynecomastia; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease; prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome; fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis; inherited, metabolic, endocrine, and toxic myopathies; myasthenia gravis, periodic paralysis; mental disorders including mood, anxiety, and schizophrenic disorders; seasonal affective disorder (SAD); akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, and Tourette's disorder; a cardiovascular disorder, such as arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular tumors, and complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery, congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, and complications of cardiac transplantation, congenital lung anomalies, atelectasis, pulmonary congestion and edema, pulmonary embolism, pulmonary hemorrhage, pulmonary infarction, pulmonary hypertension, vascular sclerosis, obstructive pulmonary disease, restrictive pulmonary disease, chronic obstructive pulmonary disease, emphysema, chronic bronchitis, bronchial asthma, bronchiectasis, bacterial pneumonia, viral and mycoplasmal pneumonia, lung abscess, pulmonary tuberculosis, diffuse interstitial diseases, pneumoconioses, sarcoidosis, idiopathic pulmonary fibrosis, desquamative interstitial pneumonitis, hypersensitivity pneumonitis, pulmonary eosinophilia bronchiolitis obliterans-organizing pneumonia,

diffuse pulmonary hemorrhage syndromes, Goodpasture's syndromes, idiopathic pulmonary hemosiderosis, pulmonary involvement in collagen-vascular disorders, pulmonary alveolar proteinosis, lung tumors, inflammatory and noninflammatory pleural effusions, pneumothorax, pleural tumors, drug-induced lung disease, radiation-induced lung disease, and complications of lung transplantation; an eye disorder such as ocular hypertension and glaucoma; a disorder of cell proliferation such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia; and a cancer, including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, colon, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus.

In another embodiment, a vector capable of expressing ENZM or a fragment or derivative thereof may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of ENZM including, but not limited to, those described above.

In a further embodiment, a composition comprising a substantially purified ENZM in conjunction with a suitable pharmaceutical carrier may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of ENZM including, but not limited to, those provided above.

In still another embodiment, an agonist which modulates the activity of ENZM may be administered to a subject to treat or prevent a disorder associated with decreased expression or activity of ENZM including, but not limited to, those listed above.

In a further embodiment, an antagonist of ENZM may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of ENZM. Examples of such disorders include, but are not limited to, those autoimmune/inflammatory disorders, infectious disorders, immune deficiencies, disorders of metabolism, reproductive disorders, neurological disorders, cardiovascular disorders, eye disorders, and cell proliferative disorders, including cancer described above. In one aspect, an antibody which specifically binds ENZM may be used directly as an antagonist or indirectly as a targeting or delivery mechanism for bringing a pharmaceutical agent to cells or tissues which express ENZM.

In an additional embodiment, a vector expressing the complement of the polynucleotide encoding ENZM may be administered to a subject to treat or prevent a disorder associated with increased expression or activity of ENZM including, but not limited to, those described above.

In other embodiments, any protein, agonist, antagonist, antibody, complementary sequence, or vector embodiments may be administered in combination with other appropriate therapeutic

agents. Selection of the appropriate agents for use in combination therapy may be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents may act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

An antagonist of ENZM may be produced using methods which are generally known in the art. In particular, purified ENZM may be used to produce antibodies or to screen libraries of pharmaceutical agents to identify those which specifically bind ENZM. Antibodies to ENZM may also be generated using methods that are well known in the art. Such antibodies may include, but are not limited to, polyclonal, monoclonal, chimeric, and single chain antibodies, Fab fragments, and fragments produced by a Fab expression library. In an embodiment, neutralizing antibodies (i.e., those which inhibit dimer formation) can be used therapeutically. Single chain antibodies (e.g., from camels or llamas) may be potent enzyme inhibitors and may have application in the design of peptide mimetics, and in the development of immuno-adsorbents and biosensors (Muyldermans, S. (2001) J. Biotechnol. 74:277-302).

For the production of antibodies, various hosts including goats, rabbits, rats, mice, camels, dromedaries, llamas, humans, and others may be immunized by injection with ENZM or with any fragment or oligopeptide thereof which has immunogenic properties. Depending on the host species, various adjuvants may be used to increase immunological response. Such adjuvants include, but are not limited to, Freund's, mineral gels such as aluminum hydroxide, and surface active substances such as lysolecithin, pluronic polyols, polyanions, peptides, oil emulsions, KLH, and dinitrophenol. Among adjuvants used in humans, BCG (bacilli Calmette-Guerin) and *Corynebacterium parvum* are especially preferable.

It is preferred that the oligopeptides, peptides, or fragments used to induce antibodies to ENZM have an amino acid sequence consisting of at least about 5 amino acids, and generally will consist of at least about 10 amino acids. It is also preferable that these oligopeptides, peptides, or fragments are substantially identical to a portion of the amino acid sequence of the natural protein. Short stretches of ENZM amino acids may be fused with those of another protein, such as KLH, and antibodies to the chimeric molecule may be produced.

Monoclonal antibodies to ENZM may be prepared using any technique which provides for the production of antibody molecules by continuous cell lines in culture. These include, but are not limited to, the hybridoma technique, the human B-cell hybridoma technique, and the EBV-hybridoma technique (Kohler, G. et al. (1975) Nature 256:495-497; Kozbor, D. et al. (1985) J. Immunol. Methods 81:31-42; Cote, R.J. et al. (1983) Proc. Natl. Acad. Sci. USA 80:2026-2030; Cole, S.P. et al. (1984) Mol. Cell Biol. 62:109-120).

In addition, techniques developed for the production of "chimeric antibodies," such as the splicing of mouse antibody genes to human antibody genes to obtain a molecule with appropriate antigen specificity and biological activity, can be used (Morrison, S.L. et al. (1984) Proc. Natl. Acad. Sci. USA 81:6851-6855; Neuberger, M.S. et al. (1984) Nature 312:604-608; Takeda, S. et al. (1985) Nature 314:452-454). Alternatively, techniques described for the production of single chain antibodies may be adapted, using methods known in the art, to produce ENZM-specific single chain antibodies. Antibodies with related specificity, but of distinct idiotypic composition, may be generated by chain shuffling from random combinatorial immunoglobulin libraries (Burton, D.R. (1991) Proc. Natl. Acad. Sci. USA 88:10134-10137).

Antibodies may also be produced by inducing *in vivo* production in the lymphocyte population or by screening immunoglobulin libraries or panels of highly specific binding reagents as disclosed in the literature (Orlandi, R. et al. (1989) Proc. Natl. Acad. Sci. USA 86:3833-3837; Winter, G. et al. (1991) Nature 349:293-299).

Antibody fragments which contain specific binding sites for ENZM may also be generated. For example, such fragments include, but are not limited to, $F(ab')_2$ fragments produced by pepsin digestion of the antibody molecule and Fab fragments generated by reducing the disulfide bridges of the F(ab')2 fragments. Alternatively, Fab expression libraries may be constructed to allow rapid and easy identification of monoclonal Fab fragments with the desired specificity (Huse, W.D. et al. (1989) Science 246:1275-1281).

Various immunoassays may be used for screening to identify antibodies having the desired specificity. Numerous protocols for competitive binding or immunoradiometric assays using either polyclonal or monoclonal antibodies with established specificities are well known in the art. Such immunoassays typically involve the measurement of complex formation between ENZM and its specific antibody. A two-site, monoclonal-based immunoassay utilizing monoclonal antibodies reactive to two non-interfering ENZM epitopes is generally used, but a competitive binding assay may also be employed (Pound, *supra*).

Various methods such as Scatchard analysis in conjunction with radioimmunoassay techniques may be used to assess the affinity of antibodies for ENZM. Affinity is expressed as an association constant, K_a, which is defined as the molar concentration of ENZM-antibody complex divided by the molar concentrations of free antigen and free antibody under equilibrium conditions. The K_a determined for a preparation of polyclonal antibodies, which are heterogeneous in their affinities for multiple ENZM epitopes, represents the average affinity, or avidity, of the antibodies for ENZM. The K_a determined for a preparation of monoclonal antibodies, which are monospecific for a particular ENZM epitope, represents a true measure of affinity. High-affinity antibody preparations with K_a ranging from about 10^9 to 10^{12} L/mole are preferred for use in immunoassays in which the

ENZM-antibody complex must withstand rigorous manipulations. Low-affinity antibody preparations with K_a ranging from about 10^6 to 10^7 L/mole are preferred for use in immunopurification and similar procedures which ultimately require dissociation of ENZM, preferably in active form, from the antibody (Catty, D. (1988) Antibodies, Volume I: A Practical Approach, IRL Press, Washington DC; Liddell, J.E. and A. Cryer (1991) A Practical Guide to Monoclonal Antibodies, John Wiley & Sons, New York NY).

The titer and avidity of polyclonal antibody preparations may be further evaluated to determine the quality and suitability of such preparations for certain downstream applications. For example, a polyclonal antibody preparation containing at least 1-2 mg specific antibody/ml, preferably 5-10 mg specific antibody/ml, is generally employed in procedures requiring precipitation of ENZM-antibody complexes. Procedures for evaluating antibody specificity, titer, and avidity, and guidelines for antibody quality and usage in various applications, are generally available (Catty, *supra*; Coligan et al., *supra*).

In another embodiment of the invention, polynucleotides encoding ENZM, or any fragment or complement thereof, may be used for therapeutic purposes. In one aspect, modifications of gene expression can be achieved by designing complementary sequences or antisense molecules (DNA, RNA, PNA, or modified oligonucleotides) to the coding or regulatory regions of the gene encoding ENZM. Such technology is well known in the art, and antisense oligonucleotides or larger fragments can be designed from various locations along the coding or control regions of sequences encoding ENZM (Agrawal, S., ed. (1996) Antisense Therapeutics, Humana Press, Totawa NJ).

In therapeutic use, any gene delivery system suitable for introduction of the antisense sequences into appropriate target cells can be used. Antisense sequences can be delivered intracellularly in the form of an expression plasmid which, upon transcription, produces a sequence complementary to at least a portion of the cellular sequence encoding the target protein (Slater, J.E. et al. (1998) J. Allergy Clin. Immunol. 102:469-475; Scanlon, K.J. et al. (1995) 9:1288-1296). Antisense sequences can also be introduced intracellularly through the use of viral vectors, such as retrovirus and adeno-associated virus vectors (Miller, A.D. (1990) Blood 76:271; Ausubel et al., *supra*; Uckert, W. and W. Walther (1994) Pharmacol. Ther. 63:323-347). Other gene delivery mechanisms include liposome-derived systems, artificial viral envelopes, and other systems known in the art (Rossi, J.J. (1995) Br. Med. Bull. 51:217-225; Boado, R.J. et al. (1998) J. Pharm. Sci. 87:1308-1315; Morris, M.C. et al. (1997) Nucleic Acids Res. 25:2730-2736).

In another embodiment of the invention, polynucleotides encoding ENZM may be used for somatic or germline gene therapy. Gene therapy may be performed to (i) correct a genetic deficiency (e.g., in the cases of severe combined immunodeficiency (SCID)-X1 disease characterized by X-linked inheritance (Cavazzana-Calvo, M. et al. (2000) Science 288:669-672), severe combined

immunodeficiency syndrome associated with an inherited adenosine deaminase (ADA) deficiency (Blaese, R.M. et al. (1995) Science 270:475-480; Bordignon, C. et al. (1995) Science 270:470-475), cystic fibrosis (Zabner, J. et al. (1993) Cell 75:207-216; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:643-666; Crystal, R.G. et al. (1995) Hum. Gene Therapy 6:667-703), thalassamias, familial hypercholesterolemia, and hemophilia resulting from Factor VIII or Factor IX deficiencies (Crystal, R.G. (1995) Science 270:404-410; Verma, I.M. and N. Somia (1997) Nature 389:239-242)), (ii) express a conditionally lethal gene product (e.g., in the case of cancers which result from unregulated cell proliferation), or (iii) express a protein which affords protection against intracellular parasites (e.g., against human retroviruses, such as human immunodeficiency virus (HIV) (Baltimore, D. (1988) Nature 335:395-396; Poeschla, E. et al. (1996) Proc. Natl. Acad. Sci. USA 93:11395-11399), hepatitis B or C virus (HBV, HCV); fungal parasites, such as *Candida albicans* and *Paracoccidioides brasiliensis*; and protozoan parasites such as *Plasmodium falciparum* and *Trypanosoma cruzi*). In the case where a genetic deficiency in ENZM expression or regulation causes disease, the expression of ENZM from an appropriate population of transduced cells may alleviate the clinical manifestations caused by the genetic deficiency.

In a further embodiment of the invention, diseases or disorders caused by deficiencies in ENZM are treated by constructing mammalian expression vectors encoding ENZM and introducing these vectors by mechanical means into ENZM-deficient cells. Mechanical transfer technologies for use with cells *in vivo* or *ex vitro* include (i) direct DNA microinjection into individual cells, (ii) ballistic gold particle delivery, (iii) liposome-mediated transfection, (iv) receptor-mediated gene transfer, and (v) the use of DNA transposons (Morgan, R.A. and W.F. Anderson (1993) Annu. Rev. Biochem. 62:191-217; Ivics, Z. (1997) Cell 91:501-510; Boulay, J.-L. and H. Récipon (1998) Curr. Opin. Biotechnol. 9:445-450).

Expression vectors that may be effective for the expression of ENZM include, but are not limited to, the PCDNA 3.1, EPITAG, PRCCMV2, PREP, PVAX, PCR2-TOPOTA vectors (Invitrogen, Carlsbad CA), PCMV-SCRIPT, PCMV-TAG, PEGSH/PERV (Stratagene, La Jolla CA), and PTET-OFF, PTET-ON, PTRE2, PTRE2-LUC, PTK-HYG (Clontech, Palo Alto CA). ENZM may be expressed using (i) a constitutively active promoter, (e.g., from cytomegalovirus (CMV), Rous sarcoma virus (RSV), SV40 virus, thymidine kinase (TK), or β-actin genes), (ii) an inducible promoter (e.g., the tetracycline-regulated promoter (Gossen, M. and H. Bujard (1992) Proc. Natl. Acad. Sci. USA 89:5547-5551; Gossen, M. et al. (1995) Science 268:1766-1769; Rossi, F.M.V. and H.M. Blau (1998) Curr. Opin. Biotechnol. 9:451-456), commercially available in the T-REX plasmid (Invitrogen)); the ecdysone-inducible promoter (available in the plasmids PVGRXR and PIND; Invitrogen); the FK506/rapamycin inducible promoter; or the RU486/mifepristone inducible promoter (Rossi, F.M.V. and H.M. Blau, *supra*)), or (iii) a tissue-specific promoter or the native promoter of

the endogenous gene encoding ENZM from a normal individual.

Commercially available liposome transformation kits (e.g., the PERFECT LIPID TRANSFECTION KIT, available from Invitrogen) allow one with ordinary skill in the art to deliver polynucleotides to target cells in culture and require minimal effort to optimize experimental parameters. In the alternative, transformation is performed using the calcium phosphate method (Graham, F.L. and A.J. Eb (1973) Virology 52:456-467), or by electroporation (Neumann, E. et al. (1982) EMBO J. 1:841-845). The introduction of DNA to primary cells requires modification of these standardized mammalian transfection protocols.

In another embodiment of the invention, diseases or disorders caused by genetic defects with respect to ENZM expression are treated by constructing a retrovirus vector consisting of (i) the polynucleotide encoding ENZM under the control of an independent promoter or the retrovirus long terminal repeat (LTR) promoter, (ii) appropriate RNA packaging signals, and (iii) a Rev-responsive element (RRE) along with additional retrovirus *cis*-acting RNA sequences and coding sequences required for efficient vector propagation. Retrovirus vectors (e.g., PFB and PFBNEO) are commercially available (Stratagene) and are based on published data (Riviere, I. et al. (1995) Proc. Natl. Acad. Sci. USA 92:6733-6737), incorporated by reference herein. The vector is propagated in an appropriate vector producing cell line (VPCL) that expresses an envelope gene with a tropism for receptors on the target cells or a promiscuous envelope protein such as VSVg (Armentano, D. et al. (1987) J. Virol. 61:1647-1650; Bender, M.A. et al. (1987) J. Virol. 61:1639-1646; Adam, M.A. and A.D. Miller (1988) J. Virol. 62:3802-3806; Dull, T. et al. (1998) J. Virol. 72:8463-8471; Zufferey, R. et al. (1998) J. Virol. 72:9873-9880). U.S. Patent No. 5,910,434 to Rigg ("Method for obtaining retrovirus packaging cell lines producing high transducing efficiency retroviral supernatant") discloses a method for obtaining retrovirus packaging cell lines and is hereby incorporated by reference. Propagation of retrovirus vectors, transduction of a population of cells (e.g., $CD4^+$ T-cells), and the return of transduced cells to a patient are procedures well known to persons skilled in the art of gene therapy and have been well documented (Ranga, U. et al. (1997) J. Virol. 71:7020-7029; Bauer, G. et al. (1997) Blood 89:2259-2267; Bonyhadi, M.L. (1997) J. Virol. 71:4707-4716; Ranga, U. et al. (1998) Proc. Natl. Acad. Sci. USA 95:1201-1206; Su, L. (1997) Blood 89:2283-2290).

In an embodiment, an adenovirus-based gene therapy delivery system is used to deliver polynucleotides encoding ENZM to cells which have one or more genetic abnormalities with respect to the expression of ENZM. The construction and packaging of adenovirus-based vectors are well known to those with ordinary skill in the art. Replication defective adenovirus vectors have proven to be versatile for importing genes encoding immunoregulatory proteins into intact islets in the pancreas (Csete, M.E. et al. (1995) Transplantation 27:263-268). Potentially useful adenoviral vectors are

described in U.S. Patent No. 5,707,618 to Armentano ("Adenovirus vectors for gene therapy"), hereby incorporated by reference. For adenoviral vectors, see also Antinozzi, P.A. et al. (1999; Annu. Rev. Nutr. 19:511-544) and Verma, I.M. and N. Somia (1997; Nature 18:389:239-242).

In another embodiment, a herpes-based, gene therapy delivery system is used to deliver polynucleotides encoding ENZM to target cells which have one or more genetic abnormalities with respect to the expression of ENZM. The use of herpes simplex virus (HSV)-based vectors may be especially valuable for introducing ENZM to cells of the central nervous system, for which HSV has a tropism. The construction and packaging of herpes-based vectors are well known to those with ordinary skill in the art. A replication-competent herpes simplex virus (HSV) type 1-based vector has been used to deliver a reporter gene to the eyes of primates (Liu, X. et al. (1999) Exp. Eye Res. 169:385-395). The construction of a HSV-1 virus vector has also been disclosed in detail in U.S. Patent No. 5,804,413 to DeLuca ("Herpes simplex virus strains for gene transfer"), which is hereby incorporated by reference. U.S. Patent No. 5,804,413 teaches the use of recombinant HSV d92 which consists of a genome containing at least one exogenous gene to be transferred to a cell under the control of the appropriate promoter for purposes including human gene therapy. Also taught by this patent are the construction and use of recombinant HSV strains deleted for ICP4, ICP27 and ICP22. For HSV vectors, see also Goins, W.F. et al. (1999; J. Virol. 73:519-532) and Xu, H. et al. (1994; Dev. Biol. 163:152-161). The manipulation of cloned herpesvirus sequences, the generation of recombinant virus following the transfection of multiple plasmids containing different segments of the large herpesvirus genomes, the growth and propagation of herpesvirus, and the infection of cells with herpesvirus are techniques well known to those of ordinary skill in the art.

In another embodiment, an alphavirus (positive, single-stranded RNA virus) vector is used to deliver polynucleotides encoding ENZM to target cells. The biology of the prototypic alphavirus, Semliki Forest Virus (SFV), has been studied extensively and gene transfer vectors have been based on the SFV genome (Garoff, H. and K.-J. Li (1998) Curr. Opin. Biotechnol. 9:464-469). During alphavirus RNA replication, a subgenomic RNA is generated that normally encodes the viral capsid proteins. This subgenomic RNA replicates to higher levels than the full length genomic RNA, resulting in the overproduction of capsid proteins relative to the viral proteins with enzymatic activity (e.g., protease and polymerase). Similarly, inserting the coding sequence for ENZM into the alphavirus genome in place of the capsid-coding region results in the production of a large number of ENZM-coding RNAs and the synthesis of high levels of ENZM in vector transduced cells. While alphavirus infection is typically associated with cell lysis within a few days, the ability to establish a persistent infection in hamster normal kidney cells (BHK-21) with a variant of Sindbis virus (SIN) indicates that the lytic replication of alphaviruses can be altered to suit the needs of the gene therapy application (Dryga, S.A. et al. (1997) Virology 228:74-83). The wide host range of alphaviruses will

allow the introduction of ENZM into a variety of cell types. The specific transduction of a subset of cells in a population may require the sorting of cells prior to transduction. The methods of manipulating infectious cDNA clones of alphaviruses, performing alphavirus cDNA and RNA transfections, and performing alphavirus infections, are well known to those with ordinary skill in the art.

Oligonucleotides derived from the transcription initiation site, e.g., between about positions -10 and +10 from the start site, may also be employed to inhibit gene expression. Similarly, inhibition can be achieved using triple helix base-pairing methodology. Triple helix pairing is useful because it causes inhibition of the ability of the double helix to open sufficiently for the binding of polymerases, transcription factors, or regulatory molecules. Recent therapeutic advances using triplex DNA have been described in the literature (Gee, J.E. et al. (1994) in Huber, B.E. and B.I. Carr, Molecular and Immunologic Approaches, Futura Publishing, Mt. Kisco NY, pp. 163-177). A complementary sequence or antisense molecule may also be designed to block translation of mRNA by preventing the transcript from binding to ribosomes.

Ribozymes, enzymatic RNA molecules, may also be used to catalyze the specific cleavage of RNA. The mechanism of ribozyme action involves sequence-specific hybridization of the ribozyme molecule to complementary target RNA, followed by endonucleolytic cleavage. For example, engineered hammerhead motif ribozyme molecules may specifically and efficiently catalyze endonucleolytic cleavage of RNA molecules encoding ENZM.

Specific ribozyme cleavage sites within any potential RNA target are initially identified by scanning the target molecule for ribozyme cleavage sites, including the following sequences: GUA, GUU, and GUC. Once identified, short RNA sequences of between 15 and 20 ribonucleotides, corresponding to the region of the target gene containing the cleavage site, may be evaluated for secondary structural features which may render the oligonucleotide inoperable. The suitability of candidate targets may also be evaluated by testing accessibility to hybridization with complementary oligonucleotides using ribonuclease protection assays.

Complementary ribonucleic acid molecules and ribozymes may be prepared by any method known in the art for the synthesis of nucleic acid molecules. These include techniques for chemically synthesizing oligonucleotides such as solid phase phosphoramidite chemical synthesis. Alternatively, RNA molecules may be generated by *in vitro* and *in vivo* transcription of DNA molecules encoding ENZM. Such DNA sequences may be incorporated into a wide variety of vectors with suitable RNA polymerase promoters such as T7 or SP6. Alternatively, these cDNA constructs that synthesize complementary RNA, constitutively or inducibly, can be introduced into cell lines, cells, or tissues.

RNA molecules may be modified to increase intracellular stability and half-life. Possible modifications include, but are not limited to, the addition of flanking sequences at the 5' and/or 3'

ends of the molecule, or the use of phosphorothioate or 2' O-methyl rather than phosphodiesterase linkages within the backbone of the molecule. This concept is inherent in the production of PNAs and can be extended in all of these molecules by the inclusion of nontraditional bases such as inosine, queosine, and wybutosine, as well as acetyl-, methyl-, thio-, and similarly modified forms of adenine, cytidine, guanine, thymine, and uridine which are not as easily recognized by endogenous endonucleases.

In other embodiments of the invention, the expression of one or more selected polynucleotides of the present invention can be altered, inhibited, decreased, or silenced using RNA interference (RNAi) or post-transcriptional gene silencing (PTGS) methods known in the art. RNAi is a post-transcriptional mode of gene silencing in which double-stranded RNA (dsRNA) introduced into a targeted cell specifically suppresses the expression of the homologous gene (i.e., the gene bearing the sequence complementary to the dsRNA). This effectively knocks out or substantially reduces the expression of the targeted gene. PTGS can also be accomplished by use of DNA or DNA fragments as well. RNAi methods are described by Fire, A. et al. (1998; Nature 391:806-811) and Gura, T. (2000; Nature 404:804-808). PTGS can also be initiated by introduction of a complementary segment of DNA into the selected tissue using gene delivery and/or viral vector delivery methods described herein or known in the art.

RNAi can be induced in mammalian cells by the use of small interfering RNA also known as siRNA. SiRNA are shorter segments of dsRNA (typically about 21 to 23 nucleotides in length) that result *in vivo* from cleavage of introduced dsRNA by the action of an endogenous ribonuclease. SiRNA appear to be the mediators of the RNAi effect in mammals. The most effective siRNAs appear to be 21 nucleotide dsRNAs with 2 nucleotide 3' overhangs. The use of siRNA for inducing RNAi in mammalian cells is described by Elbashir, S.M. et al. (2001; Nature 411:494-498).

SiRNA can either be generated indirectly by introduction of dsRNA into the targeted cell, or directly by mammalian transfection methods and agents described herein or known in the art (such as liposome-mediated transfection, viral vector methods, or other polynucleotide delivery/introductory methods). Suitable SiRNAs can be selected by examining a transcript of the target polynucleotide (e.g., mRNA) for nucleotide sequences downstream from the AUG start codon and recording the occurrence of each nucleotide and the 3' adjacent 19 to 23 nucleotides as potential siRNA target sites, with sequences having a 21 nucleotide length being preferred. Regions to be avoided for target siRNA sites include the 5' and 3' untranslated regions (UTRs) and regions near the start codon (within 75 bases), as these may be richer in regulatory protein binding sites. UTR-binding proteins and/or translation initiation complexes may interfere with binding of the siRNP endonuclease complex. The selected target sites for siRNA can then be compared to the appropriate genome database (e.g., human, etc.) using BLAST or other sequence comparison algorithms known in the art. Target

sequences with significant homology to other coding sequences can be eliminated from consideration. The selected SiRNAs can be produced by chemical synthesis methods known in the art or by *in vitro* transcription using commercially available methods and kits such as the SILENCER siRNA construction kit (Ambion, Austin TX).

In alternative embodiments, long-term gene silencing and/or RNAi effects can be induced in selected tissue using expression vectors that continuously express siRNA. This can be accomplished using expression vectors that are engineered to express hairpin RNAs (shRNAs) using methods known in the art (see, e.g., Brummelkamp, T.R. et al. (2002) Science 296:550-553; and Paddison, P.J. et al. (2002) Genes Dev. 16:948-958). In these and related embodiments, shRNAs can be delivered to target cells using expression vectors known in the art. An example of a suitable expression vector for delivery of siRNA is the PSILENCER1.0-U6 (circular) plasmid (Ambion). Once delivered to the target tissue, shRNAs are processed *in vivo* into siRNA-like molecules capable of carrying out gene-specific silencing.

In various embodiments, the expression levels of genes targeted by RNAi or PTGS methods can be determined by assays for mRNA and/or protein analysis. Expression levels of the mRNA of a targeted gene, can be determined by northern analysis methods using, for example, the NORTHERNMAX-GLY kit (Ambion); by microarray methods; by PCR methods; by real time PCR methods; and by other RNA/polynucleotide assays known in the art or described herein. Expression levels of the protein encoded by the targeted gene can be determined by Western analysis using standard techniques known in the art.

An additional embodiment of the invention encompasses a method for screening for a compound which is effective in altering expression of a polynucleotide encoding ENZM. Compounds which may be effective in altering expression of a specific polynucleotide may include, but are not limited to, oligonucleotides, antisense oligonucleotides, triple helix-forming oligonucleotides, transcription factors and other polypeptide transcriptional regulators, and non-macromolecular chemical entities which are capable of interacting with specific polynucleotide sequences. Effective compounds may alter polynucleotide expression by acting as either inhibitors or promoters of polynucleotide expression. Thus, in the treatment of disorders associated with increased ENZM expression or activity, a compound which specifically inhibits expression of the polynucleotide encoding ENZM may be therapeutically useful, and in the treatment of disorders associated with decreased ENZM expression or activity, a compound which specifically promotes expression of the polynucleotide encoding ENZM may be therapeutically useful.

In various embodiments, one or more test compounds may be screened for effectiveness in altering expression of a specific polynucleotide. A test compound may be obtained by any method commonly known in the art, including chemical modification of a compound known to be effective in

altering polynucleotide expression; selection from an existing, commercially-available or proprietary library of naturally-occurring or non-natural chemical compounds; rational design of a compound based on chemical and/or structural properties of the target polynucleotide; and selection from a library of chemical compounds created combinatorially or randomly. A sample comprising a polynucleotide encoding ENZM is exposed to at least one test compound thus obtained. The sample may comprise, for example, an intact or permeabilized cell, or an *in vitro* cell-free or reconstituted biochemical system. Alterations in the expression of a polynucleotide encoding ENZM are assayed by any method commonly known in the art. Typically, the expression of a specific nucleotide is detected by hybridization with a probe having a nucleotide sequence complementary to the sequence of the polynucleotide encoding ENZM. The amount of hybridization may be quantified, thus forming the basis for a comparison of the expression of the polynucleotide both with and without exposure to one or more test compounds. Detection of a change in the expression of a polynucleotide exposed to a test compound indicates that the test compound is effective in altering the expression of the polynucleotide. A screen for a compound effective in altering expression of a specific polynucleotide can be carried out, for example, using a *Schizosaccharomyces pombe* gene expression system (Atkins, D. et al. (1999) U.S. Patent No. 5,932,435; Arndt, G.M. et al. (2000) Nucleic Acids Res. 28:E15) or a human cell line such as HeLa cell (Clarke, M.L. et al. (2000) Biochem. Biophys. Res. Commun. 268:8-13). A particular embodiment of the present invention involves screening a combinatorial library of oligonucleotides (such as deoxyribonucleotides, ribonucleotides, peptide nucleic acids, and modified oligonucleotides) for antisense activity against a specific polynucleotide sequence (Bruice, T.W. et al. (1997) U.S. Patent No. 5,686,242; Bruice, T.W. et al. (2000) U.S. Patent No. 6,022,691).

Many methods for introducing vectors into cells or tissues are available and equally suitable for use *in vivo*, *in vitro*, and *ex vivo*. For *ex vivo* therapy, vectors may be introduced into stem cells taken from the patient and clonally propagated for autologous transplant back into that same patient. Delivery by transfection, by liposome injections, or by polycationic amino polymers may be achieved using methods which are well known in the art (Goldman, C.K. et al. (1997) Nat. Biotechnol. 15:462-466).

Any of the therapeutic methods described above may be applied to any subject in need of such therapy, including, for example, mammals such as humans, dogs, cats, cows, horses, rabbits, and monkeys.

An additional embodiment of the invention relates to the administration of a composition which generally comprises an active ingredient formulated with a pharmaceutically acceptable excipient. Excipients may include, for example, sugars, starches, celluloses, gums, and proteins. Various formulations are commonly known and are thoroughly discussed in the latest edition of

Remington's Pharmaceutical Sciences (Maack Publishing, Easton PA). Such compositions may consist of ENZM, antibodies to ENZM, and mimetics, agonists, antagonists, or inhibitors of ENZM.

In various embodiments, the compositions described herein, such as pharmaceutical compositions, may be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, pulmonary, transdermal, subcutaneous, intraperitoneal, intranasal, enteral, topical, sublingual, or rectal means.

Compositions for pulmonary administration may be prepared in liquid or dry powder form. These compositions are generally aerosolized immediately prior to inhalation by the patient. In the case of small molecules (e.g. traditional low molecular weight organic drugs), aerosol delivery of fast-acting formulations is well-known in the art. In the case of macromolecules (e.g. larger peptides and proteins), recent developments in the field of pulmonary delivery via the alveolar region of the lung have enabled the practical delivery of drugs such as insulin to blood circulation (see, e.g., Patton, J.S. et al., U.S. Patent No. 5,997,848). Pulmonary delivery allows administration without needle injection, and obviates the need for potentially toxic penetration enhancers.

Compositions suitable for use in the invention include compositions wherein the active ingredients are contained in an effective amount to achieve the intended purpose. The determination of an effective dose is well within the capability of those skilled in the art.

Specialized forms of compositions may be prepared for direct intracellular delivery of macromolecules comprising ENZM or fragments thereof. For example, liposome preparations containing a cell-impermeable macromolecule may promote cell fusion and intracellular delivery of the macromolecule. Alternatively, ENZM or a fragment thereof may be joined to a short cationic N-terminal portion from the HIV Tat-1 protein. Fusion proteins thus generated have been found to transduce into the cells of all tissues, including the brain, in a mouse model system (Schwarze, S.R. et al. (1999) Science 285:1569-1572).

For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays, e.g., of neoplastic cells, or in animal models such as mice, rats, rabbits, dogs, monkeys, or pigs. An animal model may also be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

A therapeutically effective dose refers to that amount of active ingredient, for example ENZM or fragments thereof, antibodies of ENZM, and agonists, antagonists or inhibitors of ENZM, which ameliorates the symptoms or condition. Therapeutic efficacy and toxicity may be determined by standard pharmaceutical procedures in cell cultures or with experimental animals, such as by calculating the ED_{50} (the dose therapeutically effective in 50% of the population) or LD_{50} (the dose lethal to 50% of the population) statistics. The dose ratio of toxic to therapeutic effects is the

therapeutic index, which can be expressed as the LD_{50}/ED_{50} ratio. Compositions which exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies are used to formulate a range of dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that includes the ED_{50} with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, the sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject requiring treatment. Dosage and administration are adjusted to provide sufficient levels of the active moiety or to maintain the desired effect. Factors which may be taken into account include the severity of the disease state, the general health of the subject, the age, weight, and gender of the subject, time and frequency of administration, drug combination(s), reaction sensitivities, and response to therapy. Long-acting compositions may be administered every 3 to 4 days, every week, or biweekly depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts may vary from about 0.1 μg to 100,000 μg, up to a total dose of about 1 gram, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

DIAGNOSTICS

In another embodiment, antibodies which specifically bind ENZM may be used for the diagnosis of disorders characterized by expression of ENZM, or in assays to monitor patients being treated with ENZM or agonists, antagonists, or inhibitors of ENZM. Antibodies useful for diagnostic purposes may be prepared in the same manner as described above for therapeutics. Diagnostic assays for ENZM include methods which utilize the antibody and a label to detect ENZM in human body fluids or in extracts of cells or tissues. The antibodies may be used with or without modification, and may be labeled by covalent or non-covalent attachment of a reporter molecule. A wide variety of reporter molecules, several of which are described above, are known in the art and may be used.

A variety of protocols for measuring ENZM, including ELISAs, RIAs, and FACS, are known in the art and provide a basis for diagnosing altered or abnormal levels of ENZM expression. Normal or standard values for ENZM expression are established by combining body fluids or cell extracts taken from normal mammalian subjects, for example, human subjects, with antibodies to ENZM under conditions suitable for complex formation. The amount of standard complex formation may be quantitated by various methods, such as photometric means. Quantities of ENZM expressed in subject, control, and disease samples from biopsied tissues are compared with the standard values.

Deviation between standard and subject values establishes the parameters for diagnosing disease.

In another embodiment of the invention, polynucleotides encoding ENZM may be used for diagnostic purposes. The polynucleotides which may be used include oligonucleotides, complementary RNA and DNA molecules, and PNAs. The polynucleotides may be used to detect and quantify gene expression in biopsied tissues in which expression of ENZM may be correlated with disease. The diagnostic assay may be used to determine absence, presence, and excess expression of ENZM, and to monitor regulation of ENZM levels during therapeutic intervention.

In one aspect, hybridization with PCR probes which are capable of detecting polynucleotides, including genomic sequences, encoding ENZM or closely related molecules may be used to identify nucleic acid sequences which encode ENZM. The specificity of the probe, whether it is made from a highly specific region, e.g., the 5' regulatory region, or from a less specific region, e.g., a conserved motif, and the stringency of the hybridization or amplification will determine whether the probe identifies only naturally occurring sequences encoding ENZM, allelic variants, or related sequences.

Probes may also be used for the detection of related sequences, and may have at least 50% sequence identity to any of the ENZM encoding sequences. The hybridization probes of the subject invention may be DNA or RNA and may be derived from the sequence of SEQ ID NO:56-110 or from genomic sequences including promoters, enhancers, and introns of the ENZM gene.

Means for producing specific hybridization probes for polynucleotides encoding ENZM include the cloning of polynucleotides encoding ENZM or ENZM derivatives into vectors for the production of mRNA probes. Such vectors are known in the art, are commercially available, and may be used to synthesize RNA probes *in vitro* by means of the addition of the appropriate RNA polymerases and the appropriate labeled nucleotides. Hybridization probes may be labeled by a variety of reporter groups, for example, by radionuclides such as ^{32}P or ^{35}S, or by enzymatic labels, such as alkaline phosphatase coupled to the probe via avidin/biotin coupling systems, and the like.

Polynucleotides encoding ENZM may be used for the diagnosis of disorders associated with expression of ENZM. Examples of such disorders include, but are not limited to, an autoimmune/inflammatory disorder such as acquired immunodeficiency syndrome (AIDS), Addison's disease, adult respiratory distress syndrome, allergies, ankylosing spondylitis, amyloidosis, anemia, asthma, atherosclerosis, autoimmune hemolytic anemia, autoimmune thyroiditis, autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy (APECED), bronchitis, cholecystitis, contact dermatitis, Crohn's disease, atopic dermatitis, dermatomyositis, diabetes mellitus, emphysema, episodic lymphopenia with lymphocytotoxins, erythroblastosis fetalis, erythema nodosum, atrophic gastritis, glomerulonephritis, Goodpasture's syndrome, gout, Graves' disease, Hashimoto's thyroiditis, hypereosinophilia, irritable bowel syndrome, multiple sclerosis, myasthenia gravis, myocardial or pericardial inflammation, osteoarthritis, osteoporosis, pancreatitis, polymyositis,

psoriasis, Reiter's syndrome, rheumatoid arthritis, scleroderma, Sjögren's syndrome, systemic anaphylaxis, systemic lupus erythematosus, systemic sclerosis, thrombocytopenic purpura, ulcerative colitis, uveitis, Werner syndrome, complications of cancer, hemodialysis, and extracorporeal circulation, and trauma; an infectious disorder such as a viral infection, e.g., caused by an adenovirus (acute respiratory disease, pneumonia), an arenavirus (lymphocytic choriomeningitis), a bunyavirus (Hantavirus), a coronavirus (pneumonia, chronic bronchitis), a hepadnavirus (hepatitis), a herpesvirus (herpes simplex virus, varicella-zoster virus, Epstein-Barr virus, cytomegalovirus), a flavivirus (yellow fever), an orthomyxovirus (influenza), a papillomavirus (cancer), a paramyxovirus (measles, mumps), a picornovirus (rhinovirus, poliovirus, coxsackie-virus), a polyomavirus (BK virus, JC virus), a poxvirus (smallpox), a reovirus (Colorado tick fever), a retrovirus (human immunodeficiency virus, human T lymphotropic virus), a rhabdovirus (rabies), a rotavirus (gastroenteritis), and a togavirus (encephalitis, rubella), and a bacterial infection, a fungal infection, a parasitic infection, a protozoal infection, and a helminthic infection; an immune deficiency, such as acquired immunodeficiency syndrome (AIDS), X-linked agammaglobinemia of Bruton, common variable immunodeficiency (CVI), DiGeorge's syndrome (thymic hypoplasia), thymic dysplasia, isolated IgA deficiency, severe combined immunodeficiency disease (SCID), immunodeficiency with thrombocytopenia and eczema (Wiskott-Aldrich syndrome), Chediak-Higashi syndrome, chronic granulomatous diseases, hereditary angioneurotic edema, and immunodeficiency associated with Cushing's disease; a disorder of metabolism such as Addison's disease, cerebrotendinous xanthomatosis, congenital adrenal hyperplasia, coumarin resistance, cystic fibrosis, diabetes, fatty hepatocirrhosis, fructose-1,6-diphosphatase deficiency, galactosemia, goiter, glucagonoma, glycogen storage diseases, hereditary fructose intolerance, hyperadrenalism, hypoadrenalism, hyperparathyroidism, hypoparathyroidism, hypercholesterolemia, hyperthyroidism, hypoglycemia, hypothyroidism, hyperlipidemia, hyperlipemia, a lipid myopathy, a lipodystrophy, a lysosomal storage disease, mannosidosis, neuraminidase deficiency, obesity, pentosuria phenylketonuria, pseudovitamin D-deficiency rickets; a reproductive disorder such as a disorder of prolactin production, infertility, including tubal disease, ovulatory defects, and endometriosis, a disruption of the estrous cycle, a disruption of the menstrual cycle, polycystic ovary syndrome, ovarian hyperstimulation syndrome, endometrial and ovarian tumors, uterine fibroids, autoimmune disorders, ectopic pregnancies, and teratogenesis, cancer of the breast, fibrocystic breast disease, and galactorrhea, disruptions of spermatogenesis, abnormal sperm physiology, cancer of the testis, cancer of the prostate, benign prostatic hyperplasia, prostatitis, Peyronie's disease, impotence, carcinoma of the male breast, and gynecomastia; a neurological disorder such as epilepsy, ischemic cerebrovascular disease, stroke, cerebral neoplasms, Alzheimer's disease, Pick's disease, Huntington's disease, dementia, Parkinson's disease and other extrapyramidal disorders, amyotrophic lateral sclerosis and

other motor neuron disorders, progressive neural muscular atrophy, retinitis pigmentosa, hereditary ataxias, multiple sclerosis and other demyelinating diseases, bacterial and viral meningitis, brain abscess, subdural empyema, epidural abscess, suppurative intracranial thrombophlebitis, myelitis and radiculitis, viral central nervous system disease; prion diseases including kuru, Creutzfeldt-Jakob disease, and Gerstmann-Straussler-Scheinker syndrome; fatal familial insomnia, nutritional and metabolic diseases of the nervous system, neurofibromatosis, tuberous sclerosis, cerebelloretinal hemangioblastomatosis, encephalotrigeminal syndrome, mental retardation and other developmental disorders of the central nervous system, cerebral palsy, neuroskeletal disorders, autonomic nervous system disorders, cranial nerve disorders, spinal cord diseases, muscular dystrophy and other neuromuscular disorders, peripheral nervous system disorders, dermatomyositis and polymyositis; inherited, metabolic, endocrine, and toxic myopathies; myasthenia gravis, periodic paralysis; mental disorders including mood, anxiety, and schizophrenic disorders; seasonal affective disorder (SAD); akathesia, amnesia, catatonia, diabetic neuropathy, tardive dyskinesia, dystonias, paranoid psychoses, postherpetic neuralgia, and Tourette's disorder; a cardiovascular disorder, such as arteriovenous fistula, atherosclerosis, hypertension, vasculitis, Raynaud's disease, aneurysms, arterial dissections, varicose veins, thrombophlebitis and phlebothrombosis, vascular tumors, and complications of thrombolysis, balloon angioplasty, vascular replacement, and coronary artery bypass graft surgery, congestive heart failure, ischemic heart disease, angina pectoris, myocardial infarction, hypertensive heart disease, degenerative valvular heart disease, calcific aortic valve stenosis, congenitally bicuspid aortic valve, mitral annular calcification, mitral valve prolapse, rheumatic fever and rheumatic heart disease, infective endocarditis, nonbacterial thrombotic endocarditis, endocarditis of systemic lupus erythematosus, carcinoid heart disease, cardiomyopathy, myocarditis, pericarditis, neoplastic heart disease, congenital heart disease, and complications of cardiac transplantation, congenital lung anomalies, atelectasis, pulmonary congestion and edema, pulmonary embolism, pulmonary hemorrhage, pulmonary infarction, pulmonary hypertension, vascular sclerosis, obstructive pulmonary disease, restrictive pulmonary disease, chronic obstructive pulmonary disease, emphysema, chronic bronchitis, bronchial asthma, bronchiectasis, bacterial pneumonia, viral and mycoplasmal pneumonia, lung abscess, pulmonary tuberculosis, diffuse interstitial diseases, pneumoconioses, sarcoidosis, idiopathic pulmonary fibrosis, desquamative interstitial pneumonitis, hypersensitivity pneumonitis, pulmonary eosinophilia bronchiolitis obliterans-organizing pneumonia, diffuse pulmonary hemorrhage syndromes, Goodpasture's syndromes, idiopathic pulmonary hemosiderosis, pulmonary involvement in collagen-vascular disorders, pulmonary alveolar proteinosis, lung tumors, inflammatory and noninflammatory pleural effusions, pneumothorax, pleural tumors, drug-induced lung disease, radiation-induced lung disease, and complications of lung transplantation; an eye disorder such as ocular hypertension and glaucoma; a disorder of cell

proliferation such as actinic keratosis, arteriosclerosis, atherosclerosis, bursitis, cirrhosis, hepatitis, mixed connective tissue disease (MCTD), myelofibrosis, paroxysmal nocturnal hemoglobinuria, polycythemia vera, psoriasis, primary thrombocythemia; and a cancer, including adenocarcinoma, leukemia, lymphoma, melanoma, myeloma, sarcoma, teratocarcinoma, and, in particular, cancers of the adrenal gland, bladder, bone, bone marrow, brain, breast, cervix, colon, gall bladder, ganglia, gastrointestinal tract, heart, kidney, liver, lung, muscle, ovary, pancreas, parathyroid, penis, prostate, salivary glands, skin, spleen, testis, thymus, thyroid, and uterus. Polynucleotides encoding ENZM may be used in Southern or northern analysis, dot blot, or other membrane-based technologies; in PCR technologies; in dipstick, pin, and multiformat ELISA-like assays; and in microarrays utilizing fluids or tissues from patients to detect altered ENZM expression. Such qualitative or quantitative methods are well known in the art.

In a particular embodiment, polynucleotides encoding ENZM may be used in assays that detect the presence of associated disorders, particularly those mentioned above. Polynucleotides complementary to sequences encoding ENZM may be labeled by standard methods and added to a fluid or tissue sample from a patient under conditions suitable for the formation of hybridization complexes. After a suitable incubation period, the sample is washed and the signal is quantified and compared with a standard value. If the amount of signal in the patient sample is significantly altered in comparison to a control sample then the presence of altered levels of polynucleotides encoding ENZM in the sample indicates the presence of the associated disorder. Such assays may also be used to evaluate the efficacy of a particular therapeutic treatment regimen in animal studies, in clinical trials, or to monitor the treatment of an individual patient.

In order to provide a basis for the diagnosis of a disorder associated with expression of ENZM, a normal or standard profile for expression is established. This may be accomplished by combining body fluids or cell extracts taken from normal subjects, either animal or human, with a sequence, or a fragment thereof, encoding ENZM, under conditions suitable for hybridization or amplification. Standard hybridization may be quantified by comparing the values obtained from normal subjects with values from an experiment in which a known amount of a substantially purified polynucleotide is used. Standard values obtained in this manner may be compared with values obtained from samples from patients who are symptomatic for a disorder. Deviation from standard values is used to establish the presence of a disorder.

Once the presence of a disorder is established and a treatment protocol is initiated, hybridization assays may be repeated on a regular basis to determine if the level of expression in the patient begins to approximate that which is observed in the normal subject. The results obtained from successive assays may be used to show the efficacy of treatment over a period ranging from several days to months.

With respect to cancer, the presence of an abnormal amount of transcript (either under- or overexpressed) in biopsied tissue from an individual may indicate a predisposition for the development of the disease, or may provide a means for detecting the disease prior to the appearance of actual clinical symptoms. A more definitive diagnosis of this type may allow health professionals to employ preventative measures or aggressive treatment earlier, thereby preventing the development or further progression of the cancer.

Additional diagnostic uses for oligonucleotides designed from the sequences encoding ENZM may involve the use of PCR. These oligomers may be chemically synthesized, generated enzymatically, or produced *in vitro*. Oligomers will preferably contain a fragment of a polynucleotide encoding ENZM, or a fragment of a polynucleotide complementary to the polynucleotide encoding ENZM, and will be employed under optimized conditions for identification of a specific gene or condition. Oligomers may also be employed under less stringent conditions for detection or quantification of closely related DNA or RNA sequences.

In a particular aspect, oligonucleotide primers derived from polynucleotides encoding ENZM may be used to detect single nucleotide polymorphisms (SNPs). SNPs are substitutions, insertions and deletions that are a frequent cause of inherited or acquired genetic disease in humans. Methods of SNP detection include, but are not limited to, single-stranded conformation polymorphism (SSCP) and fluorescent SSCP (fSSCP) methods. In SSCP, oligonucleotide primers derived from polynucleotides encoding ENZM are used to amplify DNA using the polymerase chain reaction (PCR). The DNA may be derived, for example, from diseased or normal tissue, biopsy samples, bodily fluids, and the like. SNPs in the DNA cause differences in the secondary and tertiary structures of PCR products in single-stranded form, and these differences are detectable using gel electrophoresis in non-denaturing gels. In fSCCP, the oligonucleotide primers are fluorescently labeled, which allows detection of the amplimers in high-throughput equipment such as DNA sequencing machines. Additionally, sequence database analysis methods, termed in silico SNP (isSNP), are capable of identifying polymorphisms by comparing the sequence of individual overlapping DNA fragments which assemble into a common consensus sequence. These computer-based methods filter out sequence variations due to laboratory preparation of DNA and sequencing errors using statistical models and automated analyses of DNA sequence chromatograms. In the alternative, SNPs may be detected and characterized by mass spectrometry using, for example, the high throughput MASSARRAY system (Sequenom, Inc., San Diego CA).

SNPs may be used to study the genetic basis of human disease. For example, at least 16 common SNPs have been associated with non-insulin-dependent diabetes mellitus. SNPs are also useful for examining differences in disease outcomes in monogenic disorders, such as cystic fibrosis, sickle cell anemia, or chronic granulomatous disease. For example, variants in the mannose-binding

lectin, MBL2, have been shown to be correlated with deleterious pulmonary outcomes in cystic fibrosis. SNPs also have utility in pharmacogenomics, the identification of genetic variants that influence a patient's response to a drug, such as life-threatening toxicity. For example, a variation in N-acetyl transferase is associated with a high incidence of peripheral neuropathy in response to the anti-tuberculosis drug isoniazid, while a variation in the core promoter of the ALOX5 gene results in diminished clinical response to treatment with an anti-asthma drug that targets the 5-lipoxygenase pathway. Analysis of the distribution of SNPs in different populations is useful for investigating genetic drift, mutation, recombination, and selection, as well as for tracing the origins of populations and their migrations (Taylor, J.G. et al. (2001) Trends Mol. Med. 7:507-512; Kwok, P.-Y. and Z. Gu (1999) Mol. Med. Today 5:538-543; Nowotny, P. et al. (2001) Curr. Opin. Neurobiol. 11:637-641).

Methods which may also be used to quantify the expression of ENZM include radiolabeling or biotinylating nucleotides, coamplification of a control nucleic acid, and interpolating results from standard curves (Melby, P.C. et al. (1993) J. Immunol. Methods 159:235-244; Duplaa, C. et al. (1993) Anal. Biochem. 212:229-236). The speed of quantitation of multiple samples may be accelerated by running the assay in a high-throughput format where the oligomer or polynucleotide of interest is presented in various dilutions and a spectrophotometric or colorimetric response gives rapid quantitation.

In further embodiments, oligonucleotides or longer fragments derived from any of the polynucleotides described herein may be used as elements on a microarray. The microarray can be used in transcript imaging techniques which monitor the relative expression levels of large numbers of genes simultaneously as described below. The microarray may also be used to identify genetic variants, mutations, and polymorphisms. This information may be used to determine gene function, to understand the genetic basis of a disorder, to diagnose a disorder, to monitor progression/regression of disease as a function of gene expression, and to develop and monitor the activities of therapeutic agents in the treatment of disease. In particular, this information may be used to develop a pharmacogenomic profile of a patient in order to select the most appropriate and effective treatment regimen for that patient. For example, therapeutic agents which are highly effective and display the fewest side effects may be selected for a patient based on his/her pharmacogenomic profile.

In another embodiment, ENZM, fragments of ENZM, or antibodies specific for ENZM may be used as elements on a microarray. The microarray may be used to monitor or measure protein-protein interactions, drug-target interactions, and gene expression profiles, as described above.

A particular embodiment relates to the use of the polynucleotides of the present invention to generate a transcript image of a tissue or cell type. A transcript image represents the global pattern of gene expression by a particular tissue or cell type. Global gene expression patterns are analyzed by

quantifying the number of expressed genes and their relative abundance under given conditions and at a given time (Seilhamer et al., "Comparative Gene Transcript Analysis," U.S. Patent No. 5,840,484; hereby expressly incorporated by reference herein). Thus a transcript image may be generated by hybridizing the polynucleotides of the present invention or their complements to the totality of transcripts or reverse transcripts of a particular tissue or cell type. In one embodiment, the hybridization takes place in high-throughput format, wherein the polynucleotides of the present invention or their complements comprise a subset of a plurality of elements on a microarray. The resultant transcript image would provide a profile of gene activity.

Transcript images may be generated using transcripts isolated from tissues, cell lines, biopsies, or other biological samples. The transcript image may thus reflect gene expression *in vivo*, as in the case of a tissue or biopsy sample, or *in vitro*, as in the case of a cell line.

Transcript images which profile the expression of the polynucleotides of the present invention may also be used in conjunction with *in vitro* model systems and preclinical evaluation of pharmaceuticals, as well as toxicological testing of industrial and naturally-occurring environmental compounds. All compounds induce characteristic gene expression patterns, frequently termed molecular fingerprints or toxicant signatures, which are indicative of mechanisms of action and toxicity (Nuwaysir, E.F. et al. (1999) Mol. Carcinog. 24:153-159; Steiner, S. and N.L. Anderson (2000) Toxicol. Lett. 112-113:467-471). If a test compound has a signature similar to that of a compound with known toxicity, it is likely to share those toxic properties. These fingerprints or signatures are most useful and refined when they contain expression information from a large number of genes and gene families. Ideally, a genome-wide measurement of expression provides the highest quality signature. Even genes whose expression is not altered by any tested compounds are important as well, as the levels of expression of these genes are used to normalize the rest of the expression data. The normalization procedure is useful for comparison of expression data after treatment with different compounds. While the assignment of gene function to elements of a toxicant signature aids in interpretation of toxicity mechanisms, knowledge of gene function is not necessary for the statistical matching of signatures which leads to prediction of toxicity (see, for example, Press Release 00-02 from the National Institute of Environmental Health Sciences, released February 29, 2000, available at http://www.niehs.nih.gov/oc/news/toxchip.htm). Therefore, it is important and desirable in toxicological screening using toxicant signatures to include all expressed gene sequences.

In an embodiment, the toxicity of a test compound can be assessed by treating a biological sample containing nucleic acids with the test compound. Nucleic acids that are expressed in the treated biological sample are hybridized with one or more probes specific to the polynucleotides of the present invention, so that transcript levels corresponding to the polynucleotides of the present invention may be quantified. The transcript levels in the treated biological sample are compared with

levels in an untreated biological sample. Differences in the transcript levels between the two samples are indicative of a toxic response caused by the test compound in the treated sample.

Another embodiment relates to the use of the polypeptides disclosed herein to analyze the proteome of a tissue or cell type. The term proteome refers to the global pattern of protein expression in a particular tissue or cell type. Each protein component of a proteome can be subjected individually to further analysis. Proteome expression patterns, or profiles, are analyzed by quantifying the number of expressed proteins and their relative abundance under given conditions and at a given time. A profile of a cell's proteome may thus be generated by separating and analyzing the polypeptides of a particular tissue or cell type. In one embodiment, the separation is achieved using two-dimensional gel electrophoresis, in which proteins from a sample are separated by isoelectric focusing in the first dimension, and then according to molecular weight by sodium dodecyl sulfate slab gel electrophoresis in the second dimension (Steiner and Anderson, *supra*). The proteins are visualized in the gel as discrete and uniquely positioned spots, typically by staining the gel with an agent such as Coomassie Blue or silver or fluorescent stains. The optical density of each protein spot is generally proportional to the level of the protein in the sample. The optical densities of equivalently positioned protein spots from different samples, for example, from biological samples either treated or untreated with a test compound or therapeutic agent, are compared to identify any changes in protein spot density related to the treatment. The proteins in the spots are partially sequenced using, for example, standard methods employing chemical or enzymatic cleavage followed by mass spectrometry. The identity of the protein in a spot may be determined by comparing its partial sequence, preferably of at least 5 contiguous amino acid residues, to the polypeptide sequences of interest. In some cases, further sequence data may be obtained for definitive protein identification.

A proteomic profile may also be generated using antibodies specific for ENZM to quantify the levels of ENZM expression. In one embodiment, the antibodies are used as elements on a microarray, and protein expression levels are quantified by exposing the microarray to the sample and detecting the levels of protein bound to each array element (Lueking, A. et al. (1999) Anal. Biochem. 270:103-111; Mendoze, L.G. et al. (1999) Biotechniques 27:778-788). Detection may be performed by a variety of methods known in the art, for example, by reacting the proteins in the sample with a thiol- or amino-reactive fluorescent compound and detecting the amount of fluorescence bound at each array element.

Toxicant signatures at the proteome level are also useful for toxicological screening, and should be analyzed in parallel with toxicant signatures at the transcript level. There is a poor correlation between transcript and protein abundances for some proteins in some tissues (Anderson, N.L. and J. Seilhamer (1997) Electrophoresis 18:533-537), so proteome toxicant signatures may be useful in the analysis of compounds which do not significantly affect the transcript image, but which

alter the proteomic profile. In addition, the analysis of transcripts in body fluids is difficult, due to rapid degradation of mRNA, so proteomic profiling may be more reliable and informative in such cases.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins that are expressed in the treated biological sample are separated so that the amount of each protein can be quantified. The amount of each protein is compared to the amount of the corresponding protein in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample. Individual proteins are identified by sequencing the amino acid residues of the individual proteins and comparing these partial sequences to the polypeptides of the present invention.

In another embodiment, the toxicity of a test compound is assessed by treating a biological sample containing proteins with the test compound. Proteins from the biological sample are incubated with antibodies specific to the polypeptides of the present invention. The amount of protein recognized by the antibodies is quantified. The amount of protein in the treated biological sample is compared with the amount in an untreated biological sample. A difference in the amount of protein between the two samples is indicative of a toxic response to the test compound in the treated sample.

Microarrays may be prepared, used, and analyzed using methods known in the art (Brennan, T.M. et al. (1995) U.S. Patent No. 5,474,796; Schena, M. et al. (1996) Proc. Natl. Acad. Sci. USA 93:10614-10619; Baldeschweiler et al. (1995) PCT application WO95/25116; Shalon, D. et al. (1995) PCT application WO95/35505; Heller, R.A. et al. (1997) Proc. Natl. Acad. Sci. USA 94:2150-2155; Heller, M.J. et al. (1997) U.S. Patent No. 5,605,662). Various types of microarrays are well known and thoroughly described in Schena, M., ed. (1999; DNA Microarrays: A Practical Approach, Oxford University Press, London).

In another embodiment of the invention, nucleic acid sequences encoding ENZM may be used to generate hybridization probes useful in mapping the naturally occurring genomic sequence. Either coding or noncoding sequences may be used, and in some instances, noncoding sequences may be preferable over coding sequences. For example, conservation of a coding sequence among members of a multi-gene family may potentially cause undesired cross hybridization during chromosomal mapping. The sequences may be mapped to a particular chromosome, to a specific region of a chromosome, or to artificial chromosome constructions, e.g., human artificial chromosomes (HACs), yeast artificial chromosomes (YACs), bacterial artificial chromosomes (BACs), bacterial P1 constructions, or single chromosome cDNA libraries (Harrington, J.J. et al. (1997) Nat. Genet. 15:345-355; Price, C.M. (1993) Blood Rev. 7:127-134; Trask, B.J. (1991) Trends

Genet. 7:149-154). Once mapped, the nucleic acid sequences may be used to develop genetic linkage maps, for example, which correlate the inheritance of a disease state with the inheritance of a particular chromosome region or restriction fragment length polymorphism (RFLP) (Lander, E.S. and D. Botstein (1986) Proc. Natl. Acad. Sci. USA 83:7353-7357).

Fluorescent *in situ* hybridization (FISH) may be correlated with other physical and genetic map data (Heinz-Ulrich, et al. (1995) in Meyers, *supra*, pp. 965-968). Examples of genetic map data can be found in various scientific journals or at the Online Mendelian Inheritance in Man (OMIM) World Wide Web site. Correlation between the location of the gene encoding ENZM on a physical map and a specific disorder, or a predisposition to a specific disorder, may help define the region of DNA associated with that disorder and thus may further positional cloning efforts.

In situ hybridization of chromosomal preparations and physical mapping techniques, such as linkage analysis using established chromosomal markers, may be used for extending genetic maps. Often the placement of a gene on the chromosome of another mammalian species, such as mouse, may reveal associated markers even if the exact chromosomal locus is not known. This information is valuable to investigators searching for disease genes using positional cloning or other gene discovery techniques. Once the gene or genes responsible for a disease or syndrome have been crudely localized by genetic linkage to a particular genomic region, e.g., ataxia-telangiectasia to 11q22-23, any sequences mapping to that area may represent associated or regulatory genes for further investigation (Gatti, R.A. et al. (1988) Nature 336:577-580). The nucleotide sequence of the instant invention may also be used to detect differences in the chromosomal location due to translocation, inversion, etc., among normal, carrier, or affected individuals.

In another embodiment of the invention, ENZM, its catalytic or immunogenic fragments, or oligopeptides thereof can be used for screening libraries of compounds in any of a variety of drug screening techniques. The fragment employed in such screening may be free in solution, affixed to a solid support, borne on a cell surface, or located intracellularly. The formation of binding complexes between ENZM and the agent being tested may be measured.

Another technique for drug screening provides for high throughput screening of compounds having suitable binding affinity to the protein of interest (Geysen, et al. (1984) PCT application WO84/03564). In this method, large numbers of different small test compounds are synthesized on a solid substrate. The test compounds are reacted with ENZM, or fragments thereof, and washed. Bound ENZM is then detected by methods well known in the art. Purified ENZM can also be coated directly onto plates for use in the aforementioned drug screening techniques. Alternatively, non-neutralizing antibodies can be used to capture the peptide and immobilize it on a solid support.

In another embodiment, one may use competitive drug screening assays in which neutralizing antibodies capable of binding ENZM specifically compete with a test compound for binding ENZM.

In this manner, antibodies can be used to detect the presence of any peptide which shares one or more antigenic determinants with ENZM.

In additional embodiments, the nucleotide sequences which encode ENZM may be used in any molecular biology techniques that have yet to be developed, provided the new techniques rely on properties of nucleotide sequences that are currently known, including, but not limited to, such properties as the triplet genetic code and specific base pair interactions.

Without further elaboration, it is believed that one skilled in the art can, using the preceding description, utilize the present invention to its fullest extent. The following embodiments are, therefore, to be construed as merely illustrative, and not limitative of the remainder of the disclosure in any way whatsoever.

The disclosures of all patents, applications, and publications mentioned above and below, including U.S. Ser. No. 60/368,721, and U.S. Ser. No. 60/368/799, are hereby expressly incorporated by reference.

EXAMPLES

I. Construction of cDNA Libraries

Incyte cDNAs were derived from cDNA libraries described in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA). Some tissues were homogenized and lysed in guanidinium isothiocyanate, while others were homogenized and lysed in phenol or in a suitable mixture of denaturants, such as TRIZOL (Invitrogen), a monophasic solution of phenol and guanidine isothiocyanate. The resulting lysates were centrifuged over CsCl cushions or extracted with chloroform. RNA was precipitated from the lysates with either isopropanol or sodium acetate and ethanol, or by other routine methods.

Phenol extraction and precipitation of RNA were repeated as necessary to increase RNA purity. In some cases, RNA was treated with DNase. For most libraries, poly(A)+ RNA was isolated using oligo d(T)-coupled paramagnetic particles (Promega), OLIGOTEX latex particles (QIAGEN, Chatsworth CA), or an OLIGOTEX mRNA purification kit (QIAGEN). Alternatively, RNA was isolated directly from tissue lysates using other RNA isolation kits, e.g., the POLY(A)PURE mRNA purification kit (Ambion, Austin TX).

In some cases, Stratagene was provided with RNA and constructed the corresponding cDNA libraries. Otherwise, cDNA was synthesized and cDNA libraries were constructed with the UNIZAP vector system (Stratagene) or SUPERSCRIPT plasmid system (Invitrogen), using the recommended procedures or similar methods known in the art (Ausubel et al., *supra*, ch. 5). Reverse transcription was initiated using oligo d(T) or random primers. Synthetic oligonucleotide adapters were ligated to double stranded cDNA, and the cDNA was digested with the appropriate restriction enzyme or

enzymes. For most libraries, the cDNA was size-selected (300-1000 bp) using SEPHACRYL S1000, SEPHAROSE CL2B, or SEPHAROSE CL4B column chromatography (Amersham Biosciences) or preparative agarose gel electrophoresis. cDNAs were ligated into compatible restriction enzyme sites of the polylinker of a suitable plasmid, e.g., PBLUESCRIPT plasmid (Stratagene), PSPORT1 plasmid (Invitrogen, Carlsbad CA), PCDNA2.1 plasmid (Invitrogen), PBK-CMV plasmid (Stratagene), PCR2-TOPOTA plasmid (Invitrogen), PCMV-ICIS plasmid (Stratagene), pIGEN (Incyte Genomics, Palo Alto CA), pRARE (Incyte Genomics), or pINCY (Incyte Genomics), or derivatives thereof. Recombinant plasmids were transformed into competent *E. coli* cells including XL1-Blue, XL1-BlueMRF, or SOLR from Stratagene or DH5α, DH10B, or ElectroMAX DH10B from Invitrogen.

II. Isolation of cDNA Clones

Plasmids obtained as described in Example I were recovered from host cells by *in vivo* excision using the UNIZAP vector system (Stratagene) or by cell lysis. Plasmids were purified using at least one of the following: a Magic or WIZARD Minipreps DNA purification system (Promega); an AGTC Miniprep purification kit (Edge Biosystems, Gaithersburg MD); and QIAWELL 8 Plasmid, QIAWELL 8 Plus Plasmid, QIAWELL 8 Ultra Plasmid purification systems or the R.E.A.L. PREP 96 plasmid purification kit from QIAGEN. Following precipitation, plasmids were resuspended in 0.1 ml of distilled water and stored, with or without lyophilization, at 4°C.

Alternatively, plasmid DNA was amplified from host cell lysates using direct link PCR in a high-throughput format (Rao, V.B. (1994) Anal. Biochem. 216:1-14). Host cell lysis and thermal cycling steps were carried out in a single reaction mixture. Samples were processed and stored in 384-well plates, and the concentration of amplified plasmid DNA was quantified fluorometrically using PICOGREEN dye (Molecular Probes, Eugene OR) and a FLUOROSKAN II fluorescence scanner (Labsystems Oy, Helsinki, Finland).

III. Sequencing and Analysis

Incyte cDNA recovered in plasmids as described in Example II were sequenced as follows. Sequencing reactions were processed using standard methods or high-throughput instrumentation such as the ABI CATALYST 800 (Applied Biosystems) thermal cycler or the PTC-200 thermal cycler (MJ Research) in conjunction with the HYDRA microdispenser (Robbins Scientific) or the MICROLAB 2200 (Hamilton) liquid transfer system. cDNA sequencing reactions were prepared using reagents provided by Amersham Biosciences or supplied in ABI sequencing kits such as the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems). Electrophoretic separation of cDNA sequencing reactions and detection of labeled polynucleotides were carried out using the MEGABACE 1000 DNA sequencing system (Amersham Biosciences); the ABI PRISM 373 or 377 sequencing system (Applied Biosystems) in conjunction with standard ABI protocols and base calling software; or other sequence analysis systems known in the art. Reading

frames within the cDNA sequences were identified using standard methods (Ausubel et al., *supra*, ch. 7). Some of the cDNA sequences were selected for extension using the techniques disclosed in Example VIII.

The polynucleotide sequences derived from Incyte cDNAs were validated by removing vector, linker, and poly(A) sequences and by masking ambiguous bases, using algorithms and programs based on BLAST, dynamic programming, and dinucleotide nearest neighbor analysis. The Incyte cDNA sequences or translations thereof were then queried against a selection of public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases, and BLOCKS, PRINTS, DOMO, PRODOM; PROTEOME databases with sequences from *Homo sapiens*, *Rattus norvegicus*, *Mus musculus*, *Caenorhabditis elegans*, *Saccharomyces cerevisiae*, *Schizosaccharomyces pombe*, and *Candida albicans* (Incyte Genomics, Palo Alto CA); hidden Markov model (HMM)-based protein family databases such as PFAM, INCY, and TIGRFAM (Haft, D.H. et al. (2001) Nucleic Acids Res. 29:41-43); and HMM-based protein domain databases such as SMART (Schultz, J. et al. (1998) Proc. Natl. Acad. Sci. USA 95:5857-5864; Letunic, I. et al. (2002) Nucleic Acids Res. 30:242-244). (HMM is a probabilistic approach which analyzes consensus primary structures of gene families; see, for example, Eddy, S.R. (1996) Curr. Opin. Struct. Biol. 6:361-365.) The queries were performed using programs based on BLAST, FASTA, BLIMPS, and HMMER. The Incyte cDNA sequences were assembled to produce full length polynucleotide sequences. Alternatively, GenBank cDNAs, GenBank ESTs, stitched sequences, stretched sequences, or Genscan-predicted coding sequences (see Examples IV and V) were used to extend Incyte cDNA assemblages to full length. Assembly was performed using programs based on Phred, Phrap, and Consed, and cDNA assemblages were screened for open reading frames using programs based on GeneMark, BLAST, and FASTA. The full length polynucleotide sequences were translated to derive the corresponding full length polypeptide sequences. Alternatively, a polypeptide may begin at any of the methionine residues of the full length translated polypeptide. Full length polypeptide sequences were subsequently analyzed by querying against databases such as the GenBank protein databases (genpept), SwissProt, the PROTEOME databases, BLOCKS, PRINTS, DOMO, PRODOM, Prosite, hidden Markov model (HMM)-based protein family databases such as PFAM, INCY, and TIGRFAM; and HMM-based protein domain databases such as SMART. Full length polynucleotide sequences are also analyzed using MACDNASIS PRO software (MiraiBio, Alameda CA) and LASERGENE software (DNASTAR). Polynucleotide and polypeptide sequence alignments are generated using default parameters specified by the CLUSTAL algorithm as incorporated into the MEGALIGN multisequence alignment program (DNASTAR), which also calculates the percent identity between aligned sequences.

Table 7 summarizes the tools, programs, and algorithms used for the analysis and assembly of

Incyte cDNA and full length sequences and provides applicable descriptions, references, and threshold parameters. The first column of Table 7 shows the tools, programs, and algorithms used, the second column provides brief descriptions thereof, the third column presents appropriate references, all of which are incorporated by reference herein in their entirety, and the fourth column presents, where applicable, the scores, probability values, and other parameters used to evaluate the strength of a match between two sequences (the higher the score or the lower the probability value, the greater the identity between two sequences).

The programs described above for the assembly and analysis of full length polynucleotide and polypeptide sequences were also used to identify polynucleotide sequence fragments from SEQ ID NO:56-110. Fragments from about 20 to about 4000 nucleotides which are useful in hybridization and amplification technologies are described in Table 4, column 2.

IV. Identification and Editing of Coding Sequences from Genomic DNA

Putative enzymes were initially identified by running the Genscan gene identification program against public genomic sequence databases (e.g., gbpri and gbhtg). Genscan is a general-purpose gene identification program which analyzes genomic DNA sequences from a variety of organisms (Burge, C. and S. Karlin (1997) J. Mol. Biol. 268:78-94; Burge, C. and S. Karlin (1998) Curr. Opin. Struct. Biol. 8:346-354). The program concatenates predicted exons to form an assembled cDNA sequence extending from a methionine to a stop codon. The output of Genscan is a FASTA database of polynucleotide and polypeptide sequences. The maximum range of sequence for Genscan to analyze at once was set to 30 kb. To determine which of these Genscan predicted cDNA sequences encode enzymes, the encoded polypeptides were analyzed by querying against PFAM models for enzymes. Potential enzymes were also identified by homology to Incyte cDNA sequences that had been annotated as enzymes. These selected Genscan-predicted sequences were then compared by BLAST analysis to the genpept and gbpri public databases. Where necessary, the Genscan-predicted sequences were then edited by comparison to the top BLAST hit from genpept to correct errors in the sequence predicted by Genscan, such as extra or omitted exons. BLAST analysis was also used to find any Incyte cDNA or public cDNA coverage of the Genscan-predicted sequences, thus providing evidence for transcription. When Incyte cDNA coverage was available, this information was used to correct or confirm the Genscan predicted sequence. Full length polynucleotide sequences were obtained by assembling Genscan-predicted coding sequences with Incyte cDNA sequences and/or public cDNA sequences using the assembly process described in Example III. Alternatively, full length polynucleotide sequences were derived entirely from edited or unedited Genscan-predicted coding sequences.

V. Assembly of Genomic Sequence Data with cDNA Sequence Data

"Stitched" Sequences

Partial cDNA sequences were extended with exons predicted by the Genscan gene identification program described in Example IV. Partial cDNAs assembled as described in Example III were mapped to genomic DNA and parsed into clusters containing related cDNAs and Genscan exon predictions from one or more genomic sequences. Each cluster was analyzed using an algorithm based on graph theory and dynamic programming to integrate cDNA and genomic information, generating possible splice variants that were subsequently confirmed, edited, or extended to create a full length sequence. Sequence intervals in which the entire length of the interval was present on more than one sequence in the cluster were identified, and intervals thus identified were considered to be equivalent by transitivity. For example, if an interval was present on a cDNA and two genomic sequences, then all three intervals were considered to be equivalent. This process allows unrelated but consecutive genomic sequences to be brought together, bridged by cDNA sequence. Intervals thus identified were then "stitched" together by the stitching algorithm in the order that they appear along their parent sequences to generate the longest possible sequence, as well as sequence variants. Linkages between intervals which proceed along one type of parent sequence (cDNA to cDNA or genomic sequence to genomic sequence) were given preference over linkages which change parent type (cDNA to genomic sequence). The resultant stitched sequences were translated and compared by BLAST analysis to the genpept and gbpri public databases. Incorrect exons predicted by Genscan were corrected by comparison to the top BLAST hit from genpept. Sequences were further extended with additional cDNA sequences, or by inspection of genomic DNA, when necessary.

"Stretched" Sequences

Partial DNA sequences were extended to full length with an algorithm based on BLAST analysis. First, partial cDNAs assembled as described in Example III were queried against public databases such as the GenBank primate, rodent, mammalian, vertebrate, and eukaryote databases using the BLAST program. The nearest GenBank protein homolog was then compared by BLAST analysis to either Incyte cDNA sequences or GenScan exon predicted sequences described in Example IV. A chimeric protein was generated by using the resultant high-scoring segment pairs (HSPs) to map the translated sequences onto the GenBank protein homolog. Insertions or deletions may occur in the chimeric protein with respect to the original GenBank protein homolog. The GenBank protein homolog, the chimeric protein, or both were used as probes to search for homologous genomic sequences from the public human genome databases. Partial DNA sequences were therefore "stretched" or extended by the addition of homologous genomic sequences. The resultant stretched sequences were examined to determine whether it contained a complete gene.

VI. Chromosomal Mapping of ENZM Encoding Polynucleotides

The sequences which were used to assemble SEQ ID NO:56-110 were compared with sequences from the Incyte LIFESEQ database and public domain databases using BLAST and other

implementations of the Smith-Waterman algorithm. Sequences from these databases that matched SEQ ID NO:56-110 were assembled into clusters of contiguous and overlapping sequences using assembly algorithms such as Phrap (Table 7). Radiation hybrid and genetic mapping data available from public resources such as the Stanford Human Genome Center (SHGC), Whitehead Institute for Genome Research (WIGR), and Généthon were used to determine if any of the clustered sequences had been previously mapped. Inclusion of a mapped sequence in a cluster resulted in the assignment of all sequences of that cluster, including its particular SEQ ID NO:, to that map location.

Map locations are represented by ranges, or intervals, of human chromosomes. The map position of an interval, in centiMorgans, is measured relative to the terminus of the chromosome's p-arm. (The centiMorgan (cM) is a unit of measurement based on recombination frequencies between chromosomal markers. On average, 1 cM is roughly equivalent to 1 megabase (Mb) of DNA in humans, although this can vary widely due to hot and cold spots of recombination.) The cM distances are based on genetic markers mapped by Généthon which provide boundaries for radiation hybrid markers whose sequences were included in each of the clusters. Human genome maps and other resources available to the public, such as the NCBI "GeneMap'99" World Wide Web site (http://www.ncbi.nlm.nih.gov/genemap/), can be employed to determine if previously identified disease genes map within or in proximity to the intervals indicated above.

Association of ENZM Polynucleotides with Parkinson's Disease

Several genes have been identified as showing linkage to autosomal dominant forms of Parkinson's Disease (PD). PD is a common neurodegenerative disorder causing bradykinesia, resting tremor, muscular rigidity, and postural instability. Cytoplasmic eosinophilic inclusions called Lewy bodies, and neuronal loss especially in the substantia nigra pars compacta, are pathological hallmarks of PD (Valente, E.M. et al (2001) Am. J. Hum. Genet. 68:895-900). Lewy body Parkinson disease has been thought to be a specific autosomal dominant disorder (Wakabayashi, K. et al. (1998) Acta Neuropath. 96:207-210). Juvenile parkinsonism may be a specific autosomal recessive disorder (Matsumine, H. et al. (1997) Am. J. Hum. Genet. 60: 588-596). (Online Mendelian Inheritance in Man, OMIM. Johns Hopkins University, Baltimore, MD. MIM Number: 168600: Sept. 9, 2002: . World Wide Web URL: http://www.ncbi.nlm.nih.gov/omim/)

Association of a disease with a chromosomal locus can be determined by lod score. Lod score is a statistical method used to test the linkage of two or more loci within families having a genetic disease. The lod score is the logarithm to base 10 of the odds in favor of linkage. Linkage is defined as the tendency of two genes located on the same chromosome to be inherited together through meiosis (*Genetics in Medicine*, Fifth Edition, (1991) Thompson, M.W. et al., W.B. Saunders Co. Philadelphia). A lod score of +3 or greater (1000:1 odds in favor of linkage) indicates a probability of 1 in 1000 that a particular marker was found solely by chance in affected individuals,

which is strong evidence that two genetic loci are linked.

One such gene implicated in PD is PARK3, which maps to 2p13 (Gasser, T. et al. (1998) Nature Genet. 18:262-265). A marker at chromosomal position D2S441 was found to have a lod score of 3.2 in the region of PARK3. This marker supported the disease association of PARK3 in the chromosomal interval from D2S134 to D2S286 (Gasser et al., *supra*). Markers located within chromosomal intervals D2S134 and D2S286, which map between 83.88 to 94.05 centiMorgans on the short arm of chromosome 2, were used to identify genes that map in the region between D2S134 and D2S286.

A second PD gene, implicated in early-onset recessive parkinsonism, is PARK6, located on chromosome 1 at 1p35-1p36. Several markers were obtained with lod scores greater than 3 including D1S199, D1S2732, D1S2828, D1S478, D1S2702, D1S2734, D1S2674 (Valente, E.M. et al, supra). These markers were used to determine the PD-relevant range of chromosome loci and identify sequences that map to chromosome 1 between D1S199 and D1S2885.

Restriction fragment length polymorphism (RFLP) markers shown to be near regions of DNA known as sequence-tagged sites (STS), have been mapped to NT_Contigs generated by the Human Genome Project using ePCR (Schuler, G.D. (1997) Genome Research 7: 541-550, and (1998) Trends Biotechnol. 16(11):456-459). Contigs containing regions of DNA with known disease-associated markers are therefore used to identify ENZM sequences that map to disease-associated regions of the genome.

Polynucleotides encoding ENZM were mapped to NT_Contigs. Contigs longer than 1Mb were broken into subcontigs of 1Mb length with overlaping sections of 100kb. A preliminary step used an algorithm, similar to MEGABLAST, to define the mRNA sequence /masked genomic DNA contig pairings. The cDNA/genomic pairings identified by the first algorithm were confirmed, and the ENZM polynucleotides mapped to DNA contigs, using SIM4 (Florea, L. et al. (1998) Genome Res. 8:967-974, version May 2000) which had been optimized for high throughput processing and strand assignment confidence. The SIM4 output of the mRNA sequence/genomic contig pairs was further processed to determine the correct location of the ENZM polynucleotides on the genomic contig, as well as their strand identity.

SEQ ID NO:62 was mapped to Contig NT_004610 from Genbank, version 128, covering a 14.87 Mb region of the genome that also contains PD-associated genetic markers D1S199 and D1S2885. The maximum distance between SEQ ID NO:62 and markers D1S199 and D1S2885, therefore, is 14.87 Mb. Thus, SEQ ID NO:62 is in proximity with genetic markers shown to consistently associate with PD. Therefore, in various embodiments, SEQ ID NO:62 can be used for one or more of the following: i) linkage analysis of persons and/or families to the PD disease region at 1p35-1p36, ii) diagnostic assays for PD, and iii) developing therapeutics and/or other treatments for

PD.

SEQ ID NO:63 was mapped to Contig NT_004610 from Genbank, version 128, covering a 14.87 Mb region of the genome that also contains PD-associated genetic markers D1S199 and D1S2885. The maximum distance between SEQ ID NO:63 and markers D1S199 and D1S2885, therefore, is 14.87 Mb. Thus, SEQ ID NO:63 is in proximity with genetic markers shown to consistently associate with PD. Therefore, in various embodiments, SEQ ID NO:63 can be used for one or more of the following: i) linkage analysis of persons and/or families to the PD disease region at 1p35-1p36, ii) diagnostic assays for PD, and iii) developing therapeutics and/or other treatments for PD.

Association of ENZM polynucleotides with Alzheimer's Disease

ENZM polynucleotides were mapped to NT_Contigs, available from NCBI, using the following procedures. Contigs longer than 1Mb were broken into subcontigs of 1Mb in length with overlapping sections of 100kb. A preliminary step used an algorithm, similar to MEGABLAST, to define the mRNA sequence/masked genomic DNA contig pairings. The cDNA/genomic pairings identified by the first algorithm were run through Sim4 (Florea, L. et al. (1998) Genome Res. 8:967-74, version May 2000) that had been optimized in house for high throughput and strand assignment confidence). The Sim4 output of the mRNA sequence/genomic contig pairs was further processed to determine the correct location of the ENZM polynucleotides on the genomic contig, and also their strand identity.

Loci on chromosomes that map to regions associated with particular diseases can be used as markers for these particular diseases. These markers then can be used to develop diagnostic and therapeutic tools for these diseases. For example, loci on chromosome 10 are associated with or linked to Alzheimer's disease (AD), a progressive neurodegenerative disease that represents the most common form of dementia (Ait-Ghezala, G. et al. (2002) Neurosci Lett. 325:87-90). AD can be inherited as an autosomal dominant trait. Further, genetic studies have focused on identification of genes that are potential targets for new treatments or improved diagnostics. The deposition and aggregation of β-amyloid in specific regions of the brain are key neuropathological hallmarks of AD. Insulin-degrading enzyme (IDE) can degrade β-amyloid Abraham, R. et al. (2001) Hum. Genet. 109:646-652). The IDE gene has been mapped near an AD-associated locus, 10q23-q25 (Espinosa R. 3rd et al. (1991) Cytogenet. Cell Genet. 57:184-186). Linkage analysis using IDE gene markers was performed on 1426 subjects from 435 families in which at least two family members were affected with AD.

A logarithm of the odds ratio for linkage (lod) score of over 3 indicates a probability of 1 in 1000 that a particular marker was found solely by chance in affected individuals. Significant linkage (lod score of 3.3) was reported between the polymorphic marker D10S583, located at 115.3 cM on

chromosome 10, and AD with age of onset ≥50 years (Betram, L. et al. (2000) Science 290:2302-2303). D10S583 maps 36 kb upstream of the IDE gene. Further analysis of this region, however, failed to show association of SNPs (single nucleotide polymorphisms) within the IDE gene and flanking regions with late-onset AD (LOAD), in a study of 134 Caucasian LOAD cases and 111 matched controls from the United Kingdom (Abraham, R. et al, supra). Thus, although the activity of IDE may not influence the susceptibility to LOAD, there is substantial linkage in the chromosomal region containing the IDE gene, marker D10S583, and AD. The IDE gene and D10S583 both map to contig NT_008769, which contains a region of chromosome 10 that is 9.16 Mb in size.

SEQ ID NO:67 mapped to a region of contig NT_008804 from GenBank (version 128), localizing SEQ ID NO:67 to within 9.16 Mb of the Alzheimer's disease locus on chromosome 10q. Thus, SEQ ID NO:67 is in proximity with loci shown to consistently associate with Alzheimer's disease. Therefore, in various embodiments, SEQ ID NO:67 can be used for one or more of the following: i) linkage analysis of persons and/or families to the AD disease region at 10q, ii) diagnostic assays for AD, and iii) developing therapeutics and/or other treatments for AD.

SEQ ID NO:101 mapped to a region of contig NT_008804 from GenBank (version 128), localizing SEQ ID NO:101 to within 9.16 Mb of the Alzheimer's disease locus on chromosome 10q. Thus, SEQ ID NO:101 is in proximity with loci shown to consistently associate with Alzheimer's disease. Therefore, in various embodiments, SEQ ID NO:101 can be used for one or more of the following: i) linkage analysis of persons and/or families to the AD disease region at 10q, ii) diagnostic assays for AD, and iii) developing therapeutics and/or other treatments for AD.

VII. Analysis of Polynucleotide Expression

Northern analysis is a laboratory technique used to detect the presence of a transcript of a gene and involves the hybridization of a labeled nucleotide sequence to a membrane on which RNAs from a particular cell type or tissue have been bound (Sambrook and Russell, *supra*, ch. 7; Ausubel et al., *supra*, ch. 4).

Analogous computer techniques applying BLAST were used to search for identical or related molecules in databases such as GenBank or LIFESEQ (Incyte Genomics). This analysis is much faster than multiple membrane-based hybridizations. In addition, the sensitivity of the computer search can be modified to determine whether any particular match is categorized as exact or similar. The basis of the search is the product score, which is defined as:

$$\frac{\text{BLAST Score} \times \text{Percent Identity}}{5 \times \text{minimum}\{\text{length(Seq. 1)}, \text{length(Seq. 2)}\}}$$

The product score takes into account both the degree of similarity between two sequences and the length of the sequence match. The product score is a normalized value between 0 and 100, and is

calculated as follows: the BLAST score is multiplied by the percent nucleotide identity and the product is divided by (5 times the length of the shorter of the two sequences). The BLAST score is calculated by assigning a score of +5 for every base that matches in a high-scoring segment pair (HSP), and -4 for every mismatch. Two sequences may share more than one HSP (separated by gaps). If there is more than one HSP, then the pair with the highest BLAST score is used to calculate the product score. The product score represents a balance between fractional overlap and quality in a BLAST alignment. For example, a product score of 100 is produced only for 100% identity over the entire length of the shorter of the two sequences being compared. A product score of 70 is produced either by 100% identity and 70% overlap at one end, or by 88% identity and 100% overlap at the other. A product score of 50 is produced either by 100% identity and 50% overlap at one end, or 79% identity and 100% overlap.

Alternatively, polynucleotides encoding ENZM are analyzed with respect to the tissue sources from which they were derived. For example, some full length sequences are assembled, at least in part, with overlapping Incyte cDNA sequences (see Example III). Each cDNA sequence is derived from a cDNA library constructed from a human tissue. Each human tissue is classified into one of the following organ/tissue categories: cardiovascular system; connective tissue; digestive system; embryonic structures; endocrine system; exocrine glands; genitalia, female; genitalia, male; germ cells; hemic and immune system; liver; musculoskeletal system; nervous system; pancreas; respiratory system; sense organs; skin; stomatognathic system; unclassified/mixed; or urinary tract. The number of libraries in each category is counted and divided by the total number of libraries across all categories. Similarly, each human tissue is classified into one of the following disease/condition categories: cancer, cell line, developmental, inflammation, neurological, trauma, cardiovascular, pooled, and other, and the number of libraries in each category is counted and divided by the total number of libraries across all categories. The resulting percentages reflect the tissue- and disease-specific expression of cDNA encoding ENZM. cDNA sequences and cDNA library/tissue information are found in the LIFESEQ GOLD database (Incyte Genomics, Palo Alto CA).

VIII. Extension of ENZM Encoding Polynucleotides

Full length polynucleotides are produced by extension of an appropriate fragment of the full length molecule using oligonucleotide primers designed from this fragment. One primer was synthesized to initiate 5' extension of the known fragment, and the other primer was synthesized to initiate 3' extension of the known fragment. The initial primers were designed using OLIGO 4.06 software (National Biosciences), or another appropriate program, to be about 22 to 30 nucleotides in length, to have a GC content of about 50% or more, and to anneal to the target sequence at temperatures of about 68°C to about 72°C. Any stretch of nucleotides which would result in hairpin structures and primer-primer dimerizations was avoided.

Selected human cDNA libraries were used to extend the sequence. If more than one extension was necessary or desired, additional or nested sets of primers were designed.

High fidelity amplification was obtained by PCR using methods well known in the art. PCR was performed in 96-well plates using the PTC-200 thermal cycler (MJ Research, Inc.). The reaction mix contained DNA template, 200 nmol of each primer, reaction buffer containing Mg^{2+}, $(NH_4)_2SO_4$, and 2-mercaptoethanol, Taq DNA polymerase (Amersham Biosciences), ELONGASE enzyme (Invitrogen), and Pfu DNA polymerase (Stratagene), with the following parameters for primer pair PCI A and PCI B: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C. In the alternative, the parameters for primer pair T7 and SK+ were as follows: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 57°C, 1 min; Step 4: 68°C, 2 min; Step 5: Steps 2, 3, and 4 repeated 20 times; Step 6: 68°C, 5 min; Step 7: storage at 4°C.

The concentration of DNA in each well was determined by dispensing 100 µl PICOGREEN quantitation reagent (0.25% (v/v) PICOGREEN; Molecular Probes, Eugene OR) dissolved in 1X TE and 0.5 µl of undiluted PCR product into each well of an opaque fluorimeter plate (Corning Costar, Acton MA), allowing the DNA to bind to the reagent. The plate was scanned in a Fluoroskan II (Labsystems Oy, Helsinki, Finland) to measure the fluorescence of the sample and to quantify the concentration of DNA. A 5 µl to 10 µl aliquot of the reaction mixture was analyzed by electrophoresis on a 1 % agarose gel to determine which reactions were successful in extending the sequence.

The extended nucleotides were desalted and concentrated, transferred to 384-well plates, digested with CviJI cholera virus endonuclease (Molecular Biology Research, Madison WI), and sonicated or sheared prior to religation into pUC 18 vector (Amersham Biosciences). For shotgun sequencing, the digested nucleotides were separated on low concentration (0.6 to 0.8%) agarose gels, fragments were excised, and agar digested with Agar ACE (Promega). Extended clones were religated using T4 ligase (New England Biolabs, Beverly MA) into pUC 18 vector (Amersham Biosciences), treated with Pfu DNA polymerase (Stratagene) to fill-in restriction site overhangs, and transfected into competent *E. coli* cells. Transformed cells were selected on antibiotic-containing media, and individual colonies were picked and cultured overnight at 37°C in 384-well plates in LB/2x carb liquid media.

The cells were lysed, and DNA was amplified by PCR using Taq DNA polymerase (Amersham Biosciences) and Pfu DNA polymerase (Stratagene) with the following parameters: Step 1: 94°C, 3 min; Step 2: 94°C, 15 sec; Step 3: 60°C, 1 min; Step 4: 72°C, 2 min; Step 5: steps 2, 3, and 4 repeated 29 times; Step 6: 72°C, 5 min; Step 7: storage at 4°C. DNA was quantified by PICOGREEN reagent (Molecular Probes) as described above. Samples with low DNA recoveries

were reamplified using the same conditions as described above. Samples were diluted with 20% dimethysulfoxide (1:2, v/v), and sequenced using DYENAMIC energy transfer sequencing primers and the DYENAMIC DIRECT kit (Amersham Biosciences) or the ABI PRISM BIGDYE Terminator cycle sequencing ready reaction kit (Applied Biosystems).

In like manner, full length polynucleotides are verified using the above procedure or are used to obtain 5' regulatory sequences using the above procedure along with oligonucleotides designed for such extension, and an appropriate genomic library.

IX. Identification of Single Nucleotide Polymorphisms in ENZM Encoding Polynucleotides

Common DNA sequence variants known as single nucleotide polymorphisms (SNPs) were identified in SEQ ID NO:56-110 using the LIFESEQ database (Incyte Genomics). Sequences from the same gene were clustered together and assembled as described in Example III, allowing the identification of all sequence variants in the gene. An algorithm consisting of a series of filters was used to distinguish SNPs from other sequence variants. Preliminary filters removed the majority of basecall errors by requiring a minimum Phred quality score of 15, and removed sequence alignment errors and errors resulting from improper trimming of vector sequences, chimeras, and splice variants. An automated procedure of advanced chromosome analysis analysed the original chromatogram files in the vicinity of the putative SNP. Clone error filters used statistically generated algorithms to identify errors introduced during laboratory processing, such as those caused by reverse transcriptase, polymerase, or somatic mutation. Clustering error filters used statistically generated algorithms to identify errors resulting from clustering of close homologs or pseudogenes, or due to contamination by non-human sequences. A final set of filters removed duplicates and SNPs found in immunoglobulins or T-cell receptors.

Certain SNPs were selected for further characterization by mass spectrometry using the high throughput MASSARRAY system (Sequenom, Inc.) to analyze allele frequencies at the SNP sites in four different human populations. The Caucasian population comprised 92 individuals (46 male, 46 female), including 83 from Utah, four French, three Venezualan, and two Amish individuals. The African population comprised 194 individuals (97 male, 97 female), all African Americans. The Hispanic population comprised 324 individuals (162 male, 162 female), all Mexican Hispanic. The Asian population comprised 126 individuals (64 male, 62 female) with a reported parental breakdown of 43% Chinese, 31% Japanese, 13% Korean, 5% Vietnamese, and 8% other Asian. Allele frequencies were first analyzed in the Caucasian population; in some cases those SNPs which showed no allelic variance in this population were not further tested in the other three populations.

X. Labeling and Use of Individual Hybridization Probes

Hybridization probes derived from SEQ ID NO:56-110 are employed to screen cDNAs, genomic DNAs, or mRNAs. Although the labeling of oligonucleotides, consisting of about 20 base

pairs, is specifically described, essentially the same procedure is used with larger nucleotide fragments. Oligonucleotides are designed using state-of-the-art software such as OLIGO 4.06 software (National Biosciences) and labeled by combining 50 pmol of each oligomer, 250 μCi of [γ-^{32}P] adenosine triphosphate (Amersham Biosciences), and T4 polynucleotide kinase (DuPont NEN, Boston MA). The labeled oligonucleotides are substantially purified using a SEPHADEX G-25 superfine size exclusion dextran bead column (Amersham Biosciences). An aliquot containing 10^7 counts per minute of the labeled probe is used in a typical membrane-based hybridization analysis of human genomic DNA digested with one of the following endonucleases: Ase I, Bgl II, Eco RI, Pst I, Xba I, or Pvu II (DuPont NEN).

The DNA from each digest is fractionated on a 0.7% agarose gel and transferred to nylon membranes (Nytran Plus, Schleicher & Schuell, Durham NH). Hybridization is carried out for 16 hours at 40°C. To remove nonspecific signals, blots are sequentially washed at room temperature under conditions of up to, for example, 0.1 x saline sodium citrate and 0.5% sodium dodecyl sulfate. Hybridization patterns are visualized using autoradiography or an alternative imaging means and compared.

XI. Microarrays

The linkage or synthesis of array elements upon a microarray can be achieved utilizing photolithography, piezoelectric printing (ink-jet printing; see, e.g., Baldeschweiler et al., *supra*), mechanical microspotting technologies, and derivatives thereof. The substrate in each of the aforementioned technologies should be uniform and solid with a non-porous surface (Schena, M., ed. (1999) DNA Microarrays: A Practical Approach, Oxford University Press, London). Suggested substrates include silicon, silica, glass slides, glass chips, and silicon wafers. Alternatively, a procedure analogous to a dot or slot blot may also be used to arrange and link elements to the surface of a substrate using thermal, UV, chemical, or mechanical bonding procedures. A typical array may be produced using available methods and machines well known to those of ordinary skill in the art and may contain any appropriate number of elements (Schena, M. et al. (1995) Science 270:467-470; Shalon, D. et al. (1996) Genome Res. 6:639-645; Marshall, A. and J. Hodgson (1998) Nat. Biotechnol. 16:27-31).

Full length cDNAs, Expressed Sequence Tags (ESTs), or fragments or oligomers thereof may comprise the elements of the microarray. Fragments or oligomers suitable for hybridization can be selected using software well known in the art such as LASERGENE software (DNASTAR). The array elements are hybridized with polynucleotides in a biological sample. The polynucleotides in the biological sample are conjugated to a fluorescent label or other molecular tag for ease of detection. After hybridization, nonhybridized nucleotides from the biological sample are removed, and a fluorescence scanner is used to detect hybridization at each array element. Alternatively, laser

desorbtion and mass spectrometry may be used for detection of hybridization. The degree of complementarity and the relative abundance of each polynucleotide which hybridizes to an element on the microarray may be assessed. In one embodiment, microarray preparation and usage is described in detail below.

Tissue or Cell Sample Preparation

Total RNA is isolated from tissue samples using the guanidinium thiocyanate method and poly(A)$^+$ RNA is purified using the oligo-(dT) cellulose method. Each poly(A)$^+$ RNA sample is reverse transcribed using MMLV reverse-transcriptase, 0.05 pg/μl oligo-(dT) primer (21mer), 1X first strand buffer, 0.03 units/μl RNase inhibitor, 500 μM dATP, 500 μM dGTP, 500 μM dTTP, 40 μM dCTP, 40 μM dCTP-Cy3 (BDS) or dCTP-Cy5 (Amersham Biosciences). The reverse transcription reaction is performed in a 25 ml volume containing 200 ng poly(A)$^+$ RNA with GEMBRIGHT kits (Incyte Genomics). Specific control poly(A)$^+$ RNAs are synthesized by *in vitro* transcription from non-coding yeast genomic DNA. After incubation at 37°C for 2 hr, each reaction sample (one with Cy3 and another with Cy5 labeling) is treated with 2.5 ml of 0.5M sodium hydroxide and incubated for 20 minutes at 85°C to the stop the reaction and degrade the RNA. Samples are purified using two successive CHROMA SPIN 30 gel filtration spin columns (Clontech, Palo Alto CA) and after combining, both reaction samples are ethanol precipitated using 1 ml of glycogen (1 mg/ml), 60 ml sodium acetate, and 300 ml of 100% ethanol. The sample is then dried to completion using a SpeedVAC (Savant Instruments Inc., Holbrook NY) and resuspended in 14 μl 5X SSC/0.2% SDS.

Microarray Preparation

Sequences of the present invention are used to generate array elements. Each array element is amplified from bacterial cells containing vectors with cloned cDNA inserts. PCR amplification uses primers complementary to the vector sequences flanking the cDNA insert. Array elements are amplified in thirty cycles of PCR from an initial quantity of 1-2 ng to a final quantity greater than 5 μg. Amplified array elements are then purified using SEPHACRYL-400 (Amersham Biosciences).

Purified array elements are immobilized on polymer-coated glass slides. Glass microscope slides (Corning) are cleaned by ultrasound in 0.1% SDS and acetone, with extensive distilled water washes between and after treatments. Glass slides are etched in 4% hydrofluoric acid (VWR Scientific Products Corporation (VWR), West Chester PA), washed extensively in distilled water, and coated with 0.05% aminopropyl silane (Sigma-Aldrich, St. Louis MO) in 95% ethanol. Coated slides are cured in a 110°C oven.

Array elements are applied to the coated glass substrate using a procedure described in U.S. Patent No. 5,807,522, incorporated herein by reference. 1 μl of the array element DNA, at an average concentration of 100 ng/μl, is loaded into the open capillary printing element by a high-speed robotic

apparatus. The apparatus then deposits about 5 nl of array element sample per slide.

Microarrays are UV-crosslinked using a STRATALINKER UV-crosslinker (Stratagene). Microarrays are washed at room temperature once in 0.2% SDS and three times in distilled water. Non-specific binding sites are blocked by incubation of microarrays in 0.2% casein in phosphate buffered saline (PBS) (Tropix, Inc., Bedford MA) for 30 minutes at 60°C followed by washes in 0.2% SDS and distilled water as before.

Hybridization

Hybridization reactions contain 9 μl of sample mixture consisting of 0.2 μg each of Cy3 and Cy5 labeled cDNA synthesis products in 5X SSC, 0.2% SDS hybridization buffer. The sample mixture is heated to 65°C for 5 minutes and is aliquoted onto the microarray surface and covered with an 1.8 cm^2 coverslip. The arrays are transferred to a waterproof chamber having a cavity just slightly larger than a microscope slide. The chamber is kept at 100% humidity internally by the addition of 140 μl of 5X SSC in a corner of the chamber. The chamber containing the arrays is incubated for about 6.5 hours at 60°C. The arrays are washed for 10 min at 45°C in a first wash buffer (1X SSC, 0.1% SDS), three times for 10 minutes each at 45°C in a second wash buffer (0.1X SSC), and dried.

Detection

Reporter-labeled hybridization complexes are detected with a microscope equipped with an Innova 70 mixed gas 10 W laser (Coherent, Inc., Santa Clara CA) capable of generating spectral lines at 488 nm for excitation of Cy3 and at 632 nm for excitation of Cy5. The excitation laser light is focused on the array using a 20X microscope objective (Nikon, Inc., Melville NY). The slide containing the array is placed on a computer-controlled X-Y stage on the microscope and raster-scanned past the objective. The 1.8 cm x 1.8 cm array used in the present example is scanned with a resolution of 20 micrometers.

In two separate scans, a mixed gas multiline laser excites the two fluorophores sequentially. Emitted light is split, based on wavelength, into two photomultiplier tube detectors (PMT R1477, Hamamatsu Photonics Systems, Bridgewater NJ) corresponding to the two fluorophores. Appropriate filters positioned between the array and the photomultiplier tubes are used to filter the signals. The emission maxima of the fluorophores used are 565 nm for Cy3 and 650 nm for Cy5. Each array is typically scanned twice, one scan per fluorophore using the appropriate filters at the laser source, although the apparatus is capable of recording the spectra from both fluorophores simultaneously.

The sensitivity of the scans is typically calibrated using the signal intensity generated by a cDNA control species added to the sample mixture at a known concentration. A specific location on the array contains a complementary DNA sequence, allowing the intensity of the signal at that

location to be correlated with a weight ratio of hybridizing species of 1:100,000. When two samples from different sources (e.g., representing test and control cells), each labeled with a different fluorophore, are hybridized to a single array for the purpose of identifying genes that are differentially expressed, the calibration is done by labeling samples of the calibrating cDNA with the two fluorophores and adding identical amounts of each to the hybridization mixture.

The output of the photomultiplier tube is digitized using a 12-bit RTI-835H analog-to-digital (A/D) conversion board (Analog Devices, Inc., Norwood MA) installed in an IBM-compatible PC computer. The digitized data are displayed as an image where the signal intensity is mapped using a linear 20-color transformation to a pseudocolor scale ranging from blue (low signal) to red (high signal). The data is also analyzed quantitatively. Where two different fluorophores are excited and measured simultaneously, the data are first corrected for optical crosstalk (due to overlapping emission spectra) between the fluorophores using each fluorophore's emission spectrum.

A grid is superimposed over the fluorescence signal image such that the signal from each spot is centered in each element of the grid. The fluorescence signal within each element is then integrated to obtain a numerical value corresponding to the average intensity of the signal. The software used for signal analysis is the GEMTOOLS gene expression analysis program (Incyte Genomics). Array elements that exhibit at least about a two-fold change in expression, a signal-to-background ratio of at least about 2.5, and an element spot size of at least about 40%, are considered to be differentially expressed.

Expression

SEQ ID NO:56, SEQ ID NO:57, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:65, SEQ ID NO:86, SEQ ID NO:95, SEQ ID NO:98, and SEQ ID NO:102 showed differential expression in tumorous tissue versus non-tumorous or healthy tissues, as determined by microarray analysis. Array elements that exhibited at least a two-fold change in expression and a signal intensity over 250 units, a signal-to-background ratio of a least 2.5, and an element spot size of at least 40% were identified as differentially expressed using the GEMTOOLS program (Incyte Genomics).

For example, SEQ ID NO:56, SEQ ID NO:57, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:86, and SEQ ID NO:98 showed differential expression in tumorous tissue versus normal tissue from either the same donor (SEQ ID NO: 57, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:86, and SEQ ID NO:98) or samples of colon tumor versus pooled samples of normal colon tissue (SEQ ID NO:56) as determined by microarray analysis. Therefore, in various embodiments, SEQ ID NO:56, SEQ ID NO:57, SEQ ID NO:60, SEQ ID NO:61, SEQ ID NO:86, and SEQ ID NO:98 can be used for one or more of the following: i) monitoring treatment of colon cancer, ii) diagnostic assays for colon cancer, and iii) developing therapeutics and/or other treatments for colon cancer.

In another example, SEQ ID NO:86 showed decreased expression in colon adenocarcinoma

tissue versus normal colon tissue from the same donor as determined by microarray analysis. Therefore, in various embodiments, SEQ ID NO:86 can be used for one or more of the following: i) monitoring treatment of colon cancer, ii) diagnostic assays for colon cancer, and iii) developing therapeutics and/or other treatments for colon cancer. SEQ ID NO:98 also showed decreased expression in colon cancer tissue when compared to donor-matched nonmalignant colon tissue, as determined by microarray analysis. Therefore, in various embodiments, SEQ ID NO:98 can be used for one or more of the following: i) monitoring treatment of colon cancer, ii) diagnostic assays for colon cancer, and iii) developing therapeutics and/or other treatments for colon cancer.

In another example, SEQ ID NO:65 showed differential expression in ovarian tumor tissue versus normal ovarian tissue from the same as determined by microarray analysis. Therefore, in various embodiments, SEQ ID NO:65 can be used for one or more of the following: i) monitoring treatment of ovarian cancer, ii) diagnostic assays for ovarian cancer, and iii) developing therapeutics and/or other treatments for ovarian cancer.

In another example, SEQ ID NO:95 and SEQ ID NO:102 showed differential expression in lung tumor tissue versus uninvolved lung tissue from the same donor, as determined by microarray analysis. SEQ ID NO:95 and SEQ ID NO:102 showed increased expression in lung squamous cell carcinoma tissue when compared to donor-matched nonmalignant lung tissue, as determined by microarray analysis. Therefore, in various embodiments, SEQ ID NO:95 and SEQ ID NO:102 can be used for one or more of the following: i) monitoring treatment of lung cancer, ii) diagnostic assays for lung cancer, and iii) developing therapeutics and/or other treatments for lung cancer.

SEQ ID NO:91, SEQ ID NO:92, and SEQ ID NO:95 showed differential expression in nonmalignant mammary epithelial cells versus various breast carcinoma cell lines, as determined by microarray analysis. Array elements that exhibited about at least a two-fold change in expression and a signal intensity over 250 units, a signal-to-background ratio of a least 2.5, and an element spot size of at least 40% were identified as differentially expressed using the GEMTOOLS program (Incyte Genomics).

SEQ ID NO:91, SEQ ID NO:92, and SEQ ID NO:95 showed decreased expression in breast ductal carcinoma cells and breast adenocarcinoma cells when compared to the nonmalignant cell line as determined by microarray analysis. Therefore, in various embodiments, SEQ ID NO:91, SEQ ID NO:92, and SEQ ID NO:95 can be used for one or more of the following: i) monitoring treatment of breast cancer, ii) diagnostic assays for breast cancer, and iii) developing therapeutics and/or other treatments for breast cancer.

XII. Complementary Polynucleotides

Sequences complementary to the ENZM-encoding sequences, or any parts thereof, are used to detect, decrease, or inhibit expression of naturally occurring ENZM. Although use of

oligonucleotides comprising from about 15 to 30 base pairs is described, essentially the same procedure is used with smaller or with larger sequence fragments. Appropriate oligonucleotides are designed using OLIGO 4.06 software (National Biosciences) and the coding sequence of ENZM. To inhibit transcription, a complementary oligonucleotide is designed from the most unique 5' sequence and used to prevent promoter binding to the coding sequence. To inhibit translation, a complementary oligonucleotide is designed to prevent ribosomal binding to the ENZM-encoding transcript.

XIII. Expression of ENZM

Expression and purification of ENZM is achieved using bacterial or virus-based expression systems. For expression of ENZM in bacteria, cDNA is subcloned into an appropriate vector containing an antibiotic resistance gene and an inducible promoter that directs high levels of cDNA transcription. Examples of such promoters include, but are not limited to, the *trp-lac* (*tac*) hybrid promoter and the T5 or T7 bacteriophage promoter in conjunction with the *lac* operator regulatory element. Recombinant vectors are transformed into suitable bacterial hosts, e.g., BL21(DE3). Antibiotic resistant bacteria express ENZM upon induction with isopropyl beta-D-thiogalactopyranoside (IPTG). Expression of ENZM in eukaryotic cells is achieved by infecting insect or mammalian cell lines with recombinant *Autographica californica* nuclear polyhedrosis virus (AcMNPV), commonly known as baculovirus. The nonessential polyhedrin gene of baculovirus is replaced with cDNA encoding ENZM by either homologous recombination or bacterial-mediated transposition involving transfer plasmid intermediates. Viral infectivity is maintained and the strong polyhedrin promoter drives high levels of cDNA transcription. Recombinant baculovirus is used to infect *Spodoptera frugiperda* (Sf9) insect cells in most cases, or human hepatocytes, in some cases. Infection of the latter requires additional genetic modifications to baculovirus (Engelhard, E.K. et al. (1994) Proc. Natl. Acad. Sci. USA 91:3224-3227; Sandig, V. et al. (1996) Hum. Gene Ther. 7:1937-1945).

In most expression systems, ENZM is synthesized as a fusion protein with, e.g., glutathione S-transferase (GST) or a peptide epitope tag, such as FLAG or 6-His, permitting rapid, single-step, affinity-based purification of recombinant fusion protein from crude cell lysates. GST, a 26-kilodalton enzyme from *Schistosoma japonicum*, enables the purification of fusion proteins on immobilized glutathione under conditions that maintain protein activity and antigenicity (Amersham Biosciences). Following purification, the GST moiety can be proteolytically cleaved from ENZM at specifically engineered sites. FLAG, an 8-amino acid peptide, enables immunoaffinity purification using commercially available monoclonal and polyclonal anti-FLAG antibodies (Eastman Kodak). 6-His, a stretch of six consecutive histidine residues, enables purification on metal-chelate resins (QIAGEN). Methods for protein expression and purification are discussed in Ausubel et al. (*supra*,

ch. 10 and 16). Purified ENZM obtained by these methods can be used directly in the assays shown in Examples XVII, XVIII, and XIX, where applicable.

XIV. Functi nal Assays

ENZM function is assessed by expressing the sequences encoding ENZM at physiologically elevated levels in mammalian cell culture systems. cDNA is subcloned into a mammalian expression vector containing a strong promoter that drives high levels of cDNA expression. Vectors of choice include PCMV SPORT plasmid (Invitrogen, Carlsbad CA) and PCR3.1 plasmid (Invitrogen), both of which contain the cytomegalovirus promoter. 5-10 μg of recombinant vector are transiently transfected into a human cell line, for example, an endothelial or hematopoietic cell line, using either liposome formulations or electroporation. 1-2 μg of an additional plasmid containing sequences encoding a marker protein are co-transfected. Expression of a marker protein provides a means to distinguish transfected cells from nontransfected cells and is a reliable predictor of cDNA expression from the recombinant vector. Marker proteins of choice include, e.g., Green Fluorescent Protein (GFP; Clontech), CD64, or a CD64-GFP fusion protein. Flow cytometry (FCM), an automated, laser optics-based technique, is used to identify transfected cells expressing GFP or CD64-GFP and to evaluate the apoptotic state of the cells and other cellular properties. FCM detects and quantifies the uptake of fluorescent molecules that diagnose events preceding or coincident with cell death. These events include changes in nuclear DNA content as measured by staining of DNA with propidium iodide; changes in cell size and granularity as measured by forward light scatter and 90 degree side light scatter; down-regulation of DNA synthesis as measured by decrease in bromodeoxyuridine uptake; alterations in expression of cell surface and intracellular proteins as measured by reactivity with specific antibodies; and alterations in plasma membrane composition as measured by the binding of fluorescein-conjugated Annexin V protein to the cell surface. Methods in flow cytometry are discussed in Ormerod, M.G. (1994; Flow Cytometry, Oxford, New York NY).

The influence of ENZM on gene expression can be assessed using highly purified populations of cells transfected with sequences encoding ENZM and either CD64 or CD64-GFP. CD64 and CD64-GFP are expressed on the surface of transfected cells and bind to conserved regions of human immunoglobulin G (IgG). Transfected cells are efficiently separated from nontransfected cells using magnetic beads coated with either human IgG or antibody against CD64 (DYNAL, Lake Success NY). mRNA can be purified from the cells using methods well known by those of skill in the art. Expression of mRNA encoding ENZM and other genes of interest can be analyzed by northern analysis or microarray techniques.

XV. Production of ENZM Specific Antibodies

ENZM substantially purified using polyacrylamide gel electrophoresis (PAGE; see, e.g., Harrington, M.G. (1990) Methods Enzymol. 182:488-495), or other purification techniques, is used to

immunize animals (e.g., rabbits, mice, etc.) and to produce antibodies using standard protocols.

Alternatively, the ENZM amino acid sequence is analyzed using LASERGENE software (DNASTAR) to determine regions of high immunogenicity, and a corresponding oligopeptide is synthesized and used to raise antibodies by means known to those of skill in the art. Methods for selection of appropriate epitopes, such as those near the C-terminus or in hydrophilic regions are well described in the art (Ausubel et al., *supra*, ch. 11).

Typically, oligopeptides of about 15 residues in length are synthesized using an ABI 431A peptide synthesizer (Applied Biosystems) using FMOC chemistry and coupled to KLH (Sigma-Aldrich, St. Louis MO) by reaction with N-maleimidobenzoyl-N-hydroxysuccinimide ester (MBS) to increase immunogenicity (Ausubel et al., *supra*). Rabbits are immunized with the oligopeptide-KLH complex in complete Freund's adjuvant. Resulting antisera are tested for antipeptide and anti-ENZM activity by, for example, binding the peptide or ENZM to a substrate, blocking with 1% BSA, reacting with rabbit antisera, washing, and reacting with radio-iodinated goat anti-rabbit IgG.

XVI. Purification of Naturally Occurring ENZM Using Specific Antibodies

Naturally occurring or recombinant ENZM is substantially purified by immunoaffinity chromatography using antibodies specific for ENZM. An immunoaffinity column is constructed by covalently coupling anti-ENZM antibody to an activated chromatographic resin, such as CNBr-activated SEPHAROSE (Amersham Biosciences). After the coupling, the resin is blocked and washed according to the manufacturer's instructions.

Media containing ENZM are passed over the immunoaffinity column, and the column is washed under conditions that allow the preferential absorbance of ENZM (e.g., high ionic strength buffers in the presence of detergent). The column is eluted under conditions that disrupt antibody/ENZM binding (e.g., a buffer of pH 2 to pH 3, or a high concentration of a chaotrope, such as urea or thiocyanate ion), and ENZM is collected.

XVII. Identification of Molecules Which Interact with ENZM

ENZM, or biologically active fragments thereof, are labeled with ^{125}I Bolton-Hunter reagent (Bolton, A.E. and W.M. Hunter (1973) Biochem. J. 133:529-539). Candidate molecules previously arrayed in the wells of a multi-well plate are incubated with the labeled ENZM, washed, and any wells with labeled ENZM complex are assayed. Data obtained using different concentrations of ENZM are used to calculate values for the number, affinity, and association of ENZM with the candidate molecules.

Alternatively, molecules interacting with ENZM are analyzed using the yeast two-hybrid system as described in Fields, S. and O. Song (1989; Nature 340:245-246), or using commercially available kits based on the two-hybrid system, such as the MATCHMAKER system (Clontech).

ENZM may also be used in the PATHCALLING process (CuraGen Corp., New Haven CT)

which employs the yeast two-hybrid system in a high-throughput manner to determine all interactions between the proteins encoded by two large libraries of genes (Nandabalan, K. et al. (2000) U.S. Patent No. 6,057,101).

XVIII. Demonstration of ENZM Activity

ENZM activity is demonstrated through a variety of specific enzyme assays; some of which are outlined below.

ENZM oxidoreductase activity is measured by the increase in extinction coefficient of NAD(P)H coenzyme at 340 nm for the measurement of oxidation activity, or the decrease in extinction coefficient of NAD(P)H coenzyme at 340 nm for the measurement of reduction activity (Dalziel, K. (1963) J. Biol. Chem. 238:2850-2858). One of three substrates may be used: Asn-βGal, biocytidine, or ubiquinone-10. The respective subunits of the enzyme reaction, for example, cytochrome c_1-b oxidoreductase and cytochrome c, are reconstituted. The reaction mixture contains a)1-2 mg/ml ENZM; and b) 15 mM substrate, 2.4 mM $NAD(P)^+$ in 0.1 M phosphate buffer, pH 7.1 (oxidation reaction), or 2.0 mM NAD(P)H, in 0.1 M Na_2HPO_4 buffer, pH 7.4 (reduction reaction); in a total volume of 0.1 ml. Changes in absorbance at 340 nm (A_{340}) are measured at 23.5°C using a recording spectrophotometer (Shimadzu Scientific Instruments, Inc., Pleasanton, CA). The amount of NAD(P)H is stoichiometrically equivalent to the amount of substrate initially present, and the change in A_{340} is a direct measure of the amount of NAD(P)H produced; $\Delta A_{340} = 6620[NADH]$. ENZM activity is proportional to the amount of NAD(P)H present in the assay.

Aldo/keto reductase activity of ENZM is proportional to the decrease in absorbance at 340 nm as NADPH is consumed (or increased absorbance if NADPH is produced, i.e., if the reverse reaction is monitored). A standard reaction mixture is 135 mM sodium phosphate buffer (pH 6.2-7.2 depending on enzyme), 0.2 mM NADPH, 0.3 M lithium sulfate, 0.5-2.5 mg ENZM and an appropriate level of substrate. The reaction is incubated at 30 °C and the reaction is monitored continuously with a spectrophotometer. ENZM activity is calculated as mol NADPH consumed / mg of ENZM.

Acyl-CoA dehydrogenase activity of ENZM is measured using an anaerobic electron transferring flavoprotein (ETF) assay. The reaction mixture comprises 50 mM Tris-HCl (pH 8.0), 0.5% glucose, and 50 μM acyl-CoA substrate (i.e., isovaleryl-CoA) that is pre-warmed to 32 °C. The mixture is depleted of oxygen by repeated exposure to vacuum followed by layering with argon. Trace amounts of oxygen are removed by the addition of glucose oxidase and catalase followed by the addition of ETF to a final concentration of 1 μM. The reaction is initiated by addition of purified ENZM or a sample containing ENZM and exciting the reaction at 342 nm. Quenching of fluorescence caused by the transfer of electrons from the substrate to ETF is monitored at 496 nm. 1 unit of acyl-CoA dehydrogenase activity is defined as the amount of ENZM required to reduce 1

μmol of ETF per minute (Reinard, T. et al. (2000) J. Biol. Chem. 275:33738-33743).

Alcohol dehydrogenase activity of ENZM is measured by following the conversion of NAD^+ to NADH at 340 nm ($\epsilon_{340} = 6.22\ mM^{-1}\ cm^{-1}$) at 25°C in 0.1 M potassium phosphate (pH 7.5), 0.1 M glycine (pH 10.0), and 2.4 mM NAD^+. Substrate (e.g., ethanol) and ENZM are then added to the reaction. The production of NADH results in an increase in absorbance at 340 nm and correlates with the oxidation of the alcohol substrate and the amount of alcohol dehydrogenase activity in the ENZM sample (Svensson, S. (1999) J. Biol. Chem. 274:29712-29719).

Aldehyde dehydrogenase activity of ENZM is measured by determining the total hydrolase + dehydrogenase activity of ENZM and subtracting the hydrolase activity. Hydrolase activity is first determined in a reaction mixture containing 0.05 M Tris-HCl (pH 7.8), 100 mM 2-mercaptoethanol, and 0.5-18 μM substrate, e.g., 10-HCO-HPteGlu (10-formyltetrahydrofolate; HPteGlu, tetrahydrofolate) or 10-FDDF (10-formyl-5,8-dideazafolate). Approximately 1μg of ENZM is added in a final volume of 1.0 ml. The reaction is monitored and read against a blank cuvette, containing all components except enzyme. The appearance of product is measured at either 295 nm for 5,8-dideazafolate or 300 nm for HPteGlu using molar extinction coefficients of $1.89x10^4$ and $2.17x10^4$ for 5,8-dideazafolate and HPteGlu, respectively. The addition of $NADP^+$ to the reaction mixture allows the measurement of both dehydrogenase and hydrolase activity (assays are performed as before). Based on the production of product in the presence of $NADP^+$ and the production of product in the absence of the cofactor, aldehyde dehydrogenase activity is calculated for ENZM. In the alternative, aldehyde dehydrogenase activity is assayed using propanal as substrate. The reaction mixture contains 60 mM sodium pyrophosphate buffer (pH 8.5), 5 mM propanal, 1 mM $NADP^+$, and ENZM in a total volume of 1 ml. Activity is determined by the increase in absorbance at 340 nm, resulting from the generation of NADPH, and is proportional to the aldehyde dehydrogenase activity in the sample (Krupenko, S.A. et al. (1995) J. Biol. Chem. 270:519-522).

6-phosphogluconate dehydrogenase activity of ENZM is measured by incubating purified ENZM, or a composition comprising ENZM, in 120 mM triethanolamine (pH 7.5), 0.1 mM EDTA, 0.5 mM $NADP^+$, and 10-150 μM 6-phosphogluconate as substrate at 20-25°C. The production of NADPH is measured fluorimetrically (340 nm excitation, 450 nm emission) and is indicative of 6-phosphogluconate dehydrogenase activity. Alternatively, the production of NADPH is measured photometrically, based on absorbance at 340 nm. The molar amount of NADPH produced in the reaction is proportional to the 6-phosphogluconate dehydrogenase activity in the sample (Tetaud et al., *supra*).

Ribonucleotide diphosphate reductase activity of ENZM is determined by incubating purified ENZM, or a composition comprising ENZM, along with dithiothreitol, Mg^{++}, and ADP, GDP, CDP, or UDP substrate. The product of the reaction, the corresponding deoxyribonucleotide, is separated

from the substrate by thin-layer chromatography. The reaction products can be distinguished from the reactants based on rates of migration. The use of radiolabeled substrates is an alternative for increasing the sensitivity of the assay. The amount of deoxyribonucleotides produced in the reaction is proportional to the amount of ribonucleotide diphosphate reductase activity in the sample (note that this is true only for pre-steady state kinetic analysis of ribonucleotide diphosphate reductase activity, as the enzyme is subject to negative feedback inhibition by products) (Nutter and Cheng, *supra*).

Dihydrodiol dehydrogenase activity of ENZM is measured by incubating purified ENZM, or a composition comprising ENZM, in a reaction mixture comprising 50 mM glycine (pH 9.0), 2.3 mM $NADP^+$, 8% DMSO, and a trans-dihydrodiol substrate, selected from the group including but not limited to, (±)-trans-naphthalene-1,2-dihydrodiol, (±)-trans-phenanthrene-1,2-dihydrodiol, and (±)-trans-chrysene-1,2-dihydrodiol. The oxidation reaction is monitored at 340 nm to detect the formation of NADPH, which is indicative of the oxidation of the substrate. The reaction mixture can also be analyzed before and after the addition of ENZM by circular dichroism to determine the stereochemistry of the reaction components and determine which enantiomers of a racemic substrate composition are oxidized by the ENZM (Penning, *supra*).

Glutathione S-transferase (GST) activity of ENZM is determined by measuring the ENZM catalyzed conjugation of GSH with 1-chloro-2,4-dinitrobenzene (CDNB), a common substrate for most GSTs. ENZM is incubated with 1 mM CDNB and 2.5 mM GSH together in 0.1M potassium phosphate buffer, pH 6.5, at 25 °C. The conjugation reaction is measured by the change in absorbance at 340 nm using an ultraviolet spectrophometer. ENZM activity is proportional to the change in absorbance at 340 nm.

15-oxoprostaglandin 13-reductase (PGR) activity of ENZM is measured following the separation of contaminating 15-hydroxyprostaglandin dehydrogenase (15-PGDH) activity by DEAE chromatography. Following isolation of PGR containing fractions (or using the purified ENZM), activity is assayed in a reaction comprising 0.1 M sodium phosphate (pH 7.4), 1 mM 2-mercaptoethanol, 20 μg substrate (e.g., 15-oxo derivatives of prostaglandins PGE_1, PGE_2, and $PGE_{2\alpha}$), and 1 mM NADH (or a higher concentration of NADPH). ENZM is added to the reaction which is then incubated for 10 min at 37 °C before termination by the addition of 0.25 ml 2 N NaOH. The amount of 15-oxo compound remaining in the sample is determined by measuring the maximum absorption at 500 nm of the terminated reaction and comparing this value to that of a terminated control reaction that received no ENZM. 1 unit of enzyme is defined as the amount required to catalyze the oxidation of 1 μmol substrate per minute and is proportional to the amount of PGR activity in the sample.

Choline dehydrogenase activity of ENZM is identified by the ability of *E. coli*, transformed with an ENZM expression vector, to grow on media containing choline as the sole carbon and

nitrogen source. The ability of the transformed bacteria to thrive is indicative of choline dehydrogenase activity (Magne Østerås, M. (1998) Proc. Natl. Acad. Sci. USA 95:11394-11399).

ENZM thioredoxin activity is assayed as described (Luthman, M. (1982) Biochemistry 21:6628-6633). Thioredoxins catalyze the formation of disulfide bonds and regulate the redox environment in cells to enable the necessary thiol:disulfide exchanges. One way to measure the thiol:disulfide exchange is by measuring the reduction of insulin in a mixture containing 0.1 M potassium phosphate, pH 7.0, 2 mM EDTA, 0.16 μM insulin, 0.33 mM DTT, and 0.48 mM NADPH. Different concentrations of ENZM are added to the mixture, and the reaction rate is followed by monitoring the oxidation of NADPH at 340 nM.

ENZM transferase activity is measured through assays such as a methyl transferase assay in which the transfer of radiolabeled methyl groups between a donor substrate and an acceptor substrate is measured (Bokar, J.A. et al. (1994) J. Biol. Chem. 269:17697-17704). Reaction mixtures (50 μl final volume) contain 15 mM HEPES, pH 7.9, 1.5 mM $MgCl_2$, 10 mM dithiothreitol, 3% polyvinylalcohol, 1.5 μCi [*methyl*-^{3}H]AdoMet (0.375 μM AdoMet) (DuPont-NEN), 0.6 μg ENZM, and acceptor substrate (0.4 μg [^{35}S]RNA or 6-mercaptopurine (6-MP) to 1 mM final concentration). Reaction mixtures are incubated at 30 °C for 30 minutes, then at 65 °C for 5 minutes. The products are separated by chromatography or electrophoresis and the level of methyl transferase activity is determined by quantification of *methyl*-^{3}H recovery.

Aminotransferase activity of ENZM is assayed by incubating samples containing ENZM for 1 hour at 37°C in the presence of 1 mM L-kynurenine and 1 mM 2-oxoglutarate in a final volume of 200 μl of 150 mM Tris acetate buffer (pH 8.0) containing 70 μM PLP. The formation of kynurenic acid is quantified by HPLC with spectrophotometric detection at 330 nm using the appropriate standards and controls well known to those skilled in the art. In the alternative, L-3-hydroxykynurenine is used as substrate and the production of xanthurenic acid is determined by HPLC analysis of the products with UV detection at 340 nm. The production of kynurenic acid and xanthurenic acid, respectively, is indicative of aminotransferase activity (Buchli et al., *supra*).

In another alternative, aminotransferase activity of ENZM is measured by determining the activity of purified ENZM or crude samples containing ENZM toward various amino and oxo acid substrates under single turnover conditions by monitoring the changes in the UV/VIS absorption spectrum of the enzyme-bound cofactor, pyridoxal 5'-phosphate (PLP). The reactions are performed at 25°C in 50 mM 4-methylmorpholine (pH 7.5) containing 9 μM purified ENZM or ENZM containing samples and substrate to be tested (amino and oxo acid substrates). The half-reaction from amino acid to oxo acid is followed by measuring the decrease in absorbance at 360 nm and the increase in absorbance at 330 nm due to the conversion of enzyme-bound PLP to pyridoxamine 5' phosphate (PMP). The specificity and relative activity of ENZM is determined by the activity of the

enzyme preparation against specific substrates (Vacca, *supra*).

ENZM chitinase activity is determined with the fluorogenic substrates 4-methylumbelliferyl chitotriose, methylumbelliferyl chitobiose, or methylumbelliferyl N-acetylglucosamine. Purified ENZM is incubated with 0.5uM substrate at pH 4.0 (0.1M citrate buffer), pH 5.0 (0.1M phosphate buffer), or pH 6.0 (0.1M Tris-HCL). After various times of incubation, the reaction is stopped by the addition of 0.1M glycine buffer, pH 10.4, and the concentration of free methylumbelliferone is determined fluorometrically. Chitinase B from *Serratia marcescens* may be used as a positive control (Hakala, *supra*).

ENZM isomerase activity is determined by measuring 2-hydroxyhepta-2,4-diene,1,7 dioate isomerase (HHDD isomerase) activity, as described by Garrido-Peritierra, A. and R.A. Cooper (1981; Eur. J. Biochem. 17:581-584). The sample is combined with 5-carboxymethyl-2-oxo-hex-3-ene-1,5, dioate (CMHD), which is the substrate for HHDD isomerase. CMHD concentration is monitored by measuring its absorbance at 246 nm. Decrease in absorbance at 246 nm is proportional to HHDD isomerase activity of ENZM.

ENZM isomerase activity such as peptidyl prolyl *cis/trans* isomerase activity can be assayed by an enzyme assay described by Rahfeld (*supra*). The assay is performed at 10°C in 35 mM HEPES buffer, pH 7.8, containing chymotrypsin (0.5 mg/ml) and ENZM at a variety of concentrations. Under these assay conditions, the substrate, Suc-Ala-Xaa-Pro-Phe-4-NA, is in equilibrium with respect to the prolyl bond, with 80-95% in *trans* and 5-20% in *cis* conformation. An aliquot (2 μl) of the substrate dissolved in dimethyl sulfoxide (10 mg/ml) is added to the reaction mixture described above. Only the *cis* isomer is a substrate for cleavage by chymotrypsin. Thus, as the substrate is isomerized by ENZM, the product is cleaved by chymotrypsin to produce 4-nitroanilide, which is detected by its absorbance at 390 nm. 4-Nitroanilide appears in a time-dependent and a ENZM concentration-dependent manner.

Alternatively, peptidyl prolyl *cis-trans* isomerase activity of ENZM can be assayed using a chromogenic peptide in a coupled assay with chymotrypsin (Fischer, G. et al. (1984) Biomed. Biochim. Acta 43:1101-1111).

UDP glucuronyltransferase activity of ENZM is measured using a colorimetric determination of free amine groups (Gibson, G.G. and P. Skett (1994) Introduction to Drug Metabolism, Blackie Academic and Professional, London). An amine-containing substrate, such as 2-aminophenol, is incubated at 37°C with an aliquot of the enzyme in a reaction buffer containing the necessary cofactors (40 mM Tris pH 8.0, 7.5 mM $MgCl_2$, 0.025% Triton X-100, 1 mM ascorbic acid, 0.75 mM UDP-glucuronic acid). After sufficient time, the reaction is stopped by addition of ice-cold 20% trichloroacetic acid in 0.1 M phosphate buffer pH 2.7, incubated on ice, and centrifuged to clarify the supernatant. Any unreacted 2-aminophenol is destroyed in this step. Sufficient freshly-prepared

sodium nitrite is then added; this step allows formation of the diazonium salt of the glucuronidated product. Excess nitrite is removed by addition of sufficient ammonium sulfamate, and the diazonium salt is reacted with an aromatic amine (for example, N-naphthylethylene diamine) to produce a colored azo compound which can be assayed spectrophotometrically (at 540 nm, for example). A standard curve can be constructed using known concentrations of aniline, which will form a chromophore with similar properties to 2-aminophenol glucuronide.

Adenylosuccinate synthetase activity of ENZM is measured by synthesis of AMP from IMP. The sample is combined with AMP. IMP concentration is monitored spectrophotometrically at 248 nm at 23°C (Wang, W. et al. (1995) J. Biol. Chem. 270:13160-13163). The increase in IMP concentration is proportional to ENZM activity.

Alternatively, AMP binding activity of ENZM is measured by combining the sample with ^{32}P-labeled AMP. The reaction is incubated at 37°C and terminated by addition of trichloroacetic acid. The acid extract is neutralized and subjected to gel electrophoresis to remove unbound label. The radioactivity retained in the gel is proportional to ENZM activity.

In another alternative, xenobiotic carboxylic acid:CoA ligase activity of ENZM is measured by combining the sample with γ-^{33}P-ATP and measuring the formation of γ-^{33}P- pyrophosphate with time (Vessey, D.A. et al. (1998) J. Biochem. Mol. Toxicol. 12:151-155).

Protein phosphatase (PP) activity can be measured by the hydrolysis of P-nitrophenyl phosphate (PNPP). ENZM is incubated together with PNPP in HEPES buffer pH 7.5, in the presence of 0.1% β-mercaptoethanol at 37°C for 60 min. The reaction is stopped by the addition of 6 ml of 10 N NaOH (Diamond, R.H. et al. (1994) Mol. Cell. Biol. 14:3752-62).

Alternatively, acid phosphatase activity of ENZM is demonstrated by incubating ENZM containing extract with 100 µl of 10 mM PNPP in 0.1 M sodium citrate, pH 4.5, and 50 µl of 40 mM NaCl at 37°C for 20 min. The reaction is stopped by the addition of 0.5 ml of 0.4 M glycine/NaOH, pH 10.4 (Saftig, P. et al. (1997) J. Biol. Chem. 272:18628-18635). The increase in light absorbance at 410 nm resulting from the hydrolysis of PNPP is measured using a spectrophotometer. The increase in light absorbance is proportional to the activity of ENZM in the assay.

In the alternative, ENZM activity is determined by measuring the amount of phosphate removed from a phosphorylated protein substrate. Reactions are performed with 2 or 4 nM ENZM in a final volume of 30 µl containing 60 mM Tris, pH 7.6, 1 mM EDTA, 1 mM EGTA, 0.1% 2-mercaptoethanol and 10 µM substrate, ^{32}P-labeled on serine/threonine or tyrosine, as appropriate. Reactions are initiated with substrate and incubated at 30° C for 10-15 min. Reactions are quenched with 450 µl of 4% (w/v) activated charcoal in 0.6 M HCl, 90 mM $Na_4P_2O_7$, and 2 mM NaH_2PO_4, then centrifuged at 12,000 × g for 5 min. Acid-soluble ^{32}Pi is quantified by liquid scintillation counting (Sinclair, C. et al. (1999) J. Biol. Chem. 274:23666-23672).

The adenosine deaminase activity of ENZM is determined by measuring the rate of deamination that occurs when adenosine substrate is incubated with ENZM. Reactions are performed with a predetermined amount of ENZM in a final volume of 3.0 ml containing 53.3 mM potassium phosphate and 0.045 mM adenosine. Assay reagents excluding ENZM are mixed in a quartz cuvette and equilibrated to 25° C. Reactions are initiated by the addition of ENZM and are mixed immediately by inversion. The decrease in light absorbance at 265 nm resulting from the hydrolysis of adenosine to inosine is measured using a spectrophotometer. The decrease in the $A_{265\ nm}$ is recorded for approximately 5 minutes. The decrease in light absorbance is proportional to the activity of ENZM in the assay.

ENZM hydrolase activity is measured by the hydrolysis of appropriate synthetic peptide substrates conjugated with various chromogenic molecules in which the degree of hydrolysis is quantified by spectrophotometric (or fluorometric) absorption of the released chromophore (Beynon and Bond, *supra*, pp.25-55). Peptide substrates are designed according to the category of protease activity as endopeptidase (serine, cysteine, aspartic proteases), aminopeptidase (leucine aminopeptidase), or carboxypeptidase (Carboxypeptidase A and B, procollagen C-proteinase).

An assay for carbonic anhydrase activity of ENZM uses the fluorescent pH indicator 8-hydroxypyrene-1,3,6-trisulfonate (pyranine) in combination with stopped-flow fluorometry to measure carbonic anhydrase activity (Shingles, et al. 1997, Anal. Biochem. 252:190-197). A pH 6.0 solution is mixed with a pH 8.0 solution and the initial rate of bicarbonate dehydration is measured. Addition of carbonic anhydrase to the pH 6.0 solution enables the measurement of the initial rate of activity at physiological temperatures with resolution times of 2 ms. Shingles et al. (*supra*) used this assay to resolve differences in activity and sensitivity to sulfonamides by comparing mammalian carbonic anhydrase isoforms. The fluorescent technique's sensitivity allows the determination of initial rates with a protein concentration as little as 65 ng/ml.

Decarboxylase activity of ENZM is measured as the release of CO_2 from labeled substrate. For example, ornithine decarboxylase activity of ENZM is assayed by measuring the release of CO_2 from L-[1-^{14}C]-ornithine (Reddy, S.G et al. (1996) J. Biol. Chem. 271:24945-24953). Activity is measured in 200 µl assay buffer (50 mM Tris/HCl, pH 7.5, 0.1 mM EDTA, 2 mM dithiothreitol, 5 mM NaF, 0.1% Brij35, 1 mM PMSF, 60 µM pyridoxal-5-phosphate) containing 0.5 mM L-ornithine plus 0.5 µCi L-[1-^{14}C]ornithine. The reactions are stopped after 15-30 minutes by addition of 1 M citric acid, and the $^{14}CO_2$ evolved is trapped on a paper disk filter saturated with 20 µl of 2 N NaOH. The radioactivity on the disks is determined by liquid scintillation spectography. The amount of $^{14}CO_2$ released is proportional to ornithine decarboxylase activity of ENZM.

ENZM activity of ENZM in the hydrolytic direction is performed spectroscopically by monitoring the appearance of the product (CoASH) formed by reaction of substrate (acyl-CoA) and

ENZM with 5,5'-Dithiobis(2-nitrobenzoic acid) (DTNB). The final reaction volume is 1 ml of 0.05 M potassium phosphate buffer, pH 8, containing 0.1 mM DTNB, 20 μg/ml bovine serum albumin, 10 μM of acyl-CoA of different lengths (C6-CoA, C10-CoA, C14-CoA and C18-CoA, Sigma), and ENZM. The reaction mixture is incubated at 22°C for 7 minutes. Hydrolytic activity is monitored spectrophotometrically by measuring absorbance at 412 nm (Poupon, V. et al. supra).

XIX. Identification of ENZM Agonists and Antagonists

Agonists or antagonists of ENZM activation or inhibition may be tested using the assays described in section XVIII. Agonists cause an increase in ENZM activity and antagonists cause a decrease in ENZM activity.

Various modifications and variations of the described compositions, methods, and systems of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. It will be appreciated that the invention provides novel and useful proteins, and their encoding polynucleotides, which can be used in the drug discovery process, as well as methods for using these compositions for the detection, diagnosis, and treatment of diseases and conditions. Although the invention has been described in connection with certain embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Nor should the description of such embodiments be considered exhaustive or limit the invention to the precise forms disclosed. Furthermore, elements from one embodiment can be readily recombined with elements from one or more other embodiments. Such combinations can form a number of embodiments within the scope of the invention. It is intended that the scope of the invention be defined by the following claims and their equivalents.

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Incyte Full Length Clones
7509350	1	7509350CD1	56	7509350CB1	90135231CA2
7509325	2	7509325CD1	57	7509325CB1	90127924CA2
7509337	3	7509337CD1	58	7509337CB1	
7509353	4	7509353CD1	59	7509353CB1	
7509354	5	7509354CD1	60	7509354CB1	90134671CA2
7509385	6	7509385CD1	61	7509385CB1	90134778CA2, 90134870CA2, 90134878CA2
7509216	7	7509216CD1	62	7509216CB1	
7509376	8	7509376CD1	63	7509376CB1	
7501927	9	7501927CD1	64	7501927CB1	
7503274	10	7503274CD1	65	7503274CB1	
7509104	11	7509104CD1	66	7509104CB1	90137940CA2
7509996	12	7509996CD1	67	7509996CB1	
7510030	13	7510030CD1	68	7510030CB1	95098929CA2
7510062	14	7510062CD1	69	7510062CB1	90098805CA2
7510217	15	7510217CD1	70	7510217CB1	90057342CA2, 90057418CA2, 90057434CA2
7510298	16	7510298CD1	71	7510298CB1	
7510299	17	7510299CD1	72	7510299CB1	
7510368	18	7510368CD1	73	7510368CB1	7346536CA2, 90080493CA2
7510369	19	7510369CD1	74	7510369CB1	
7510377	20	7510377CD1	75	7510377CB1	
7510026	21	7510026CD1	76	7510026CB1	

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Incyte Full Length Clones
7509168	22	7509168CD1	77	7509168CB1	90005961CA2, 90005969CA2, 90005977CA2, 90005985CA2, 90005993CA2, 90006053CA2, 90006061CA2, 90006069CA2, 90006077CA2, 90006085CA2, 90006093CA2, 90015774CA2, 90015782CA2, 90015874CA2, 90015890CA2, 90137760CA2, 90137776CA2, 90137812CA2, 90137852CA2, 95177201CA2, 95177209CA2, 95177225CA2, 95177233CA2, 95177241CA2, 95177265CA2, 95177281CA2, 95177289CA2, 95177309CA2, 95177333CA2, 95177341CA2, 95177365CA2, 95177373CA2, 95177433CA2, 95177441CA2, 95177457CA2, 95177465CA2, 95177481CA2, 95177501CA2, 95177525CA2, 95177541CA2, 95177557CA2, 95177565CA2
7500607	23	7500607CD1	78	7500607CB1	1701044CA2, 90060190CA2
7506079	24	7506079CD1	79	7506079CB1	90050863CA2, 90050887CA2
7509259	25	7509259CD1	80	7509259CB1	
7509263	26	7509263CD1	81	7509263CB1	90138339CA2
7509360	27	7509360CD1	82	7509360CB1	90138211CA2
7509394	28	7509394CD1	83	7509394CB1	90139022CA2, 90141515CA2, 90161805CA2, 90161892CA2
7581076	29	7581076CD1	84	7581076CB1	
7504551	30	7504551CD1	85	7504551CB1	2984590CA2, 5861829CA2, 8005785CA2
7500652	31	7500652CD1	86	7500652CB1	
7500900	32	7500900CD1	87	7500900CB1	
7501398	33	7501398CD1	88	7501398CB1	
7501417	34	7501417CD1	89	7501417CB1	
7501472	35	7501472CD1	90	7501472CB1	
7501489	36	7501489CD1	91	7501489CB1	4689179CA2, 5275642CA2, 95149684CA2, 95149792CA2, 95149844CA2
7501555	37	7501555CD1	92	7501555CB1	95166623CA2
7501561	38	7501561CD1	93	7501561CB1	60211842CA2, 6369074CA2, 7185081CA2
7506108	39	7506108CD1	94	7506108CB1	90053943CA2, 90054035CA2

Table 1

Incyte Project ID	Polypeptide SEQ ID NO:	Incyte Polypeptide ID	Polynucleotide SEQ ID NO:	Incyte Polynucleotide ID	Incyte Full Length Clones
7506123	40	7506123CD1	95	7506123CB1	5220877CA2
7506248	41	7506248CD1	96	7506248CB1	
7506347	42	7506347CD1	97	7506347CB1	2298531CA2
7509172	43	7509172CD1	98	7509172CB1	90127924CA2
7510421	44	7510421CD1	99	7510421CB1	6402912CA2, 6822291CA2
7504625	45	7504625CD1	100	7504625CB1	90067890CA2
7504776	46	7504776CD1	101	7504776CB1	
7504927	47	7504927CD1	102	7504927CB1	
7505010	48	7505010CD1	103	7505010CB1	
7505173	49	7505173CD1	104	7505173CB1	2453431CA2
7510061	50	7510061CD1	105	7510061CB1	
7510091	51	7510091CD1	106	7510091CB1	
7510109	52	7510109CD1	107	7510109CB1	
7510121	53	7510121CD1	108	7510121CB1	
7510797	54	7510797CD1	109	7510797CB1	
7504944	55	7504944CD1	110	7504944CB1	90018515CA2

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
1	7509350CD1	g1185452	6.5E-182	[Homo sapiens] cytochrome P450 monooxygenase CYP2J2. Wu, S. et al. (1996) J. Biol. Chem. 271:3460-3468.
		334960|CYP2J2	5.5E-183	[Homo sapiens][Oxidoreductase; Small molecule-binding protein] Cytochrome P450 2J2, catalyzes oxidation of arachidonic acid to eicosanoids, may modulate vascular hemostasis, bronchial and vascular tone, endothelial cell adhesion, and insulin and glucagon secretion, activity in heart may influence cardiac physiology. Ma, J. et al. (1999) J Biol Chem 274:17777-17788.
		438165|Cyp2j6	3.4E-135	[Mus musculus][Oxidoreductase; Small molecule-binding protein] Cytochrome P450 2J6, putative arachidonic acid epoxygenase that catalyzes oxidation of arachidonic acid to eicosanoids, may regulate mucus secretion and odorant clearance in olfactory mucosa; gene is induced by pyrazole indicating regulatory function. Ma, J. et al. (supra).
2	7509325CD1	g181326	9.2E-57	[Homo sapiens] cytochrome P-450 1. Okino, S. T. et al. (1987) J. Biol. Chem. 262:16072-16079.
	7509325CD1	703979|CYP2C8	7.7E-58	[Homo sapiens][Oxidoreductase; Transporter; Small molecule-binding protein][Endoplasmic reticulum; Cytoplasmic] Member of the cytochrome P450 IIC subfamily in the heme-binding monooxygenase superfamily that metabolizes steroids, fatty acids, and xenobiotics. Furuya, H. et al. (1991) Mol Pharmacol 40:375-382.
		584613|Cyp2c37	1.1E-40	[Mus musculus][Oxidoreductase; Transporter; Small molecule-binding protein] Cytochrome P450 2C37, metabolizes arachidonic acid to 12-hydroxyeicosatetraenoic acid, expressed in liver. Luo, G. et al. (1998) Arch Biochem Biophys 357:45-57.
3	7509337CD1	g8117859	2.6E-141	[Homo sapiens] sulfotransferase 1C1. Freimuth, R. R. et al. (2000) Genomics 65:157-165.

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
		367708|Rn.42866	9.6E-123	[Rattus norvegicus][Transferase] Protein with high similarity to phenol sulfotransferase (murine Mm.10108), which is involved in sulfo-conjugation of odorants and xenobiotics in olfactory perireceptor processes, member of the sulfotransferase family. Xiangrong, L. et al. (2000) Biochem Biophys Res Commun 27: 242-250.
		608220|Sult1a2	5.1E-88	[Mus musculus][Transferase][Cytoplasmic] Phenol sulfotransferase, involved in sulfo-conjugation of odorants and xenobiotics in olfactory perireceptor processes. Miyawaki, A. et al. (1996) Embo Journal 15:2050-2055.
4	7509353CD1	g15162117	4.8E-34	[Agrobacterium tumefaciens] AGR_pAT_bx66p. Hinkle, G. et al. Unpublished.
5	7509354CD1	g12653429	2.0E-213	[Homo sapiens] dopa decarboxylase (aromatic L-amino acid decarboxylase).
		335012|DDC	1.7E-214	[Homo sapiens][Lyase] Dopa decarboxylase (aromatic L-amino acid decarboxylase), required for the synthesis of the neurotransmitters dopamine, serotonin, and norepinephrine; may serve as a molecular marker for neuroblastoma and may be involved in Parkinson's disease. Rorsman, F. et al. (1995) Proc Natl Acad Sci U S A 92:8626-8629.
		476761|Ddc	5.5E-196	[Mus musculus][Lyase] Dopa decarboxylase (aromatic L-amino acid decarboxylase), may be involved in biogenic amine and catecholamine metabolism; human DDC may serve as a molecular marker for neuroblastoma and may be involved in Parkinson's disease. Tanaka, T. et al. (1989) Proc Natl Acad Sci U S A 86:8142-8146.
6	7509385CD1	g12653429	2.3E-236	[Homo sapiens] dopa decarboxylase (aromatic L-amino acid decarboxylase).
		476761|Ddc	4.5E-216	[Mus musculus][Lyase] Dopa decarboxylase (aromatic L-amino acid decarboxylase), may be involved in biogenic amine and catecholamine metabolism; human DDC may serve as a molecular marker for neuroblastoma and may be involved in Parkinson's disease. Tanaka, T. et al. (supra).

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
		590033|Ddc	1.2E-213	[Rattus norvegicus][Lyase] Dopa decarboxylase (aromatic L-amino acid decarboxylase), required for the synthesis of the neurotransmitters dopamine, serotonin, and norepinephrine; human DDC may serve as a molecular marker for neuroblastoma and may be involved in Parkinson's disease. Tanaka, T. et al. (supra).
7	7509216CD1	g460915	9.3E-32	[Homo sapiens] Ca^{2+}-dependent phospholipase A2. Chen, J. et al. (1994) J. Biol. Chem. 269:2365-2368.
		591297|Pla2g5	2.5E-29	[Rattus norvegicus][Hydrolase][Extracellular (excluding cell wall)] Phospholipase A2 group V, calcium-dependent member of the phospholipase family that has a role in phospholipid metabolism. Chen, J. et al. (1994) Biochim Biophys Acta 1215:115-120.
		430146|Pla2g5	5.1E-29	[Mus musculus][Hydrolase][Extracellular (excluding cell wall)] Group V phospholipase_a2, releases arachidonic acid and produces prostaglandin, involved in signal transduction mediated by arachidonic acid. Balboa, M. A. et al. (1996) J Biol Chem 271:32381-32384.
8	7509376CD1	g460915	9.3E-32	[Homo sapiens] Ca^{2+}-dependent phospholipase A2. Chen, J. et al. (supra).
		337002|PLA2G5	7.9E-33	[Homo sapiens][Hydrolase] Phospholipase A2 group V, calcium-dependent member of the phospholipase family that has roles in signal transduction and phospholipid metabolism. Chen, J. et al. (supra).
9	7501927CD1	g11275336	5.3E-214	[Homo sapiens] cytochrome P450. Hashizume, T. et al. (2001) Biochem. Biophys. Res. Commun. 280:1135-1141.

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
		692286|CYP4F12	4.3E-215	[Homo sapiens][Oxidoreductase] Cytochrome P450, oxidizes arachidonic acid to 18-hydroxyarachidonic acid, omega-side chain of 11,9-epoxymethano-PGH(2) and 9,11-diazo-15-deoxy-PGH(2), oxidizes other omega-side chains less efficiently, may function in xenobiotic biotransformation. Bylund, J. et al. (2001) Biochem Biophys Res Commun 280:892-897.
10	7503274CD1	g181292	3.2E-223	[Homo sapiens] sterol 27-hydroxylase. Cali, J. J. et al. (1991) J. Biol. Chem. 266:7774-7778.
		567832|CYP27A1	2.6E-224	[Homo sapiens][Oxidoreductase][Cytoplasmic; Mitochondrial] Sterol 27-hydroxylase, a cytochrome P450 enzyme that catalyzes sterol oxidation in biosynthesis of bile acid and vitamin D3, may function in macrophages to eliminate excess cholesterol and oxysterols; gene mutations cause cerebrotendinous xanthomatosis. Chen, W. et al. (1998) Biochemistry 37:15050-15056.
		333052|Rn.33817	3.3E-160	[Rattus norvegicus][Oxidoreductase][Cytoplasmic; Mitochondrial] Mitochondrial cytochrome P450 c27-25, a member of the cytochrome P450 monooxygenase family, catalyzes the 27-hydroxylation of cholesterol and the 25-hydroxylation of vitamin D3; gene may overlap with and expression may be regulated by Spin2b gene. Su, P. et al. (1990) DNA Cell. Biol. 9:657-667.
11	7509104CD1	g15488911	9.8E-183	[Homo sapiens] aldehyde dehydrogenase 3 family, member B1.
	7509104CD1	339678|ALDH3B1	8.0E-184	[Homo sapiens][Oxidoreductase] Aldehyde dehydrogenase 3 family member B1, a putative aldehyde dehydrogenase that likely catalyzes the conversion of aldehydes to the corresponding acids, may function to detoxify aldehydes generated by lipid peroxidation and alcohol metabolism. Hsu, L. C. et al. (1994) Gene 151:285-289.

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
		329470|Rn.9810	1.2E-92	[Rattus norvegicus][Oxidoreductase] Cytosolic aldehyde dehydrogenase class 3, oxidizes aromatic aldehydes and medium-chain aliphatic aldehyde products of lipid peroxidation, expression is induced by xenobiotics. Jones, D. E. Jr et al. (1988) Proc Natl Acad Sci U S A 85:1782-1786.
12	7509996CD1	g181342	4.7E-174	[Homo sapiens] cytochrome P450c17 (EC 1.14.99.9). Chung, B. C. et al. (1987) Proc. Natl. Acad. Sci. U.S.A. 84:407-411.
		334932|CYP17	3.8E-175	[Homo sapiens][Oxidoreductase; Transporter; Small molecule-binding protein][Endoplasmic reticulum; Microsomal fraction; Cytoplasmic] Steroid 17 alpha-hydroxylase/C17-20 lyase, a cytochrome P450 enzyme that catalyzes both 17-hydroxylase and 17,20-lyase activities in steroid hormone biosynthesis, an autoantigen in adrenocortical failure. Krohn, K. et al. (1992) Lancet 339:770-773.
		590011|Cyp17	2.0E-114	[Rattus norvegicus][Oxidoreductase] Steroid 17 alpha-hydroxylase/C17-20 lyase, a cytochrome P450 enzyme that acts in corticosteroid and androgen production, converts 17 alpha-hydroxypregnenolone and 17 alpha-hydroxyprogesterone into dehydroepiandrosterone and androstenedione. Nishihara, M. et al. (1988) Biochem Biophys Res Commun 154:151-158.
13	7510030CD1	g180503	9.8E-126	[Homo sapiens] di-N-acetylchitobiase. Fisher, K. J. et al. (1992) J. Biol. Chem. 267:19607-19616.
		340298|CTBS	8.0E-127	[Homo sapiens][Hydrolase][Lysosome/vacuole; Cytoplasmic] Chitobiase (di-N-acetylchitobiase), a lysosomal glycosidase involved in the degradation of asparagine-linked glycoproteins. Fisher, K. J. et al. (1992) Mol Cell Biol 12:1585-1591.
		704880|Ctbs	5.3E-98	[Rattus norvegicus][Hydrolase][Lysosome/vacuole; Cytoplasmic] Chitobiase (di-N-acetylchitobiase), a lysosomal glycosidase involved in the degradation of asparagine-linked glycoproteins. Kuranda, M. J. et al. (1986) J Biol Chem 261:5803-5809.
14	7510062CD1	g1036799	2.9E-229	[Homo sapiens] heparan N-deacetylase/N-sulfotransferase-2.

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
		336568|NDST2	2.4E-230	[Homo sapiens][Transferase; Hydrolase] Heparan glucosaminyl N-deacetylase/N-sulphotransferase-2, has a role in heparin biosynthesis. Humphries, D. E. et al. (1998) Biochem J 332:303-307.
		586593|Ndst2	2.2E-220	[Mus musculus][Transferase; Hydrolase][Golgi; Endoplasmic reticulum; Cytoplasmic] Heparin glucosaminyl N-deacetylase/N-sulphotransferase-2, has a role in heparin biosynthesis. Aikawa, J. et al. (2001) J Biol Chem 276:5876-5882.
15	7510217CD1	g553516	4.6E-73	[Homo sapiens] iduronate 2-sulfatase. Wilson, P. J. et al. (1990) Proc. Natl. Acad. Sci. U.S.A. 87:8531-8535.
		568210|IDS	3.7E-74	[Homo sapiens][Hydrolase][Lysosome/vacuole; Cytoplasmic] Iduronate 2-sulfatase, catalyzes the hydrolysis of 2-sulfate groups of L-iduronate 2-sulfate residues of dermatan sulfate and heparan sulfate in glycosaminoglycan catabolism; mutations in gene cause Hunter syndrome (mucopolysaccharidosis type II). Wilson, P. J. et al. (1990) Proc Natl Acad Sci U S A 87:8531-8535.
		585035|Ids	2.8E-60	[Mus musculus][Hydrolase][Lysosome/vacuole; Cytoplasmic] Iduronate 2-sulfatase, alterations of human IDS result in the X-linked lysosomal storage disease Hunter syndrome. Daniele, A. et al. (1993) Genomics 16:755-757.
16	7510298CD1	g1263132	1.8E-110	[Homo sapiens] tafazzins. Bione, S. et al. (1996) Nat. Genet. 12:385-389.
17	7510299CD1	g5748487	3.1E-186	[Homo sapiens] UDP-N-acetylglucosamine: alpha-1,3-D-mannoside beta-1,4-N-acetylglucosaminyltransferase IV. Yoshida, A. et al. (1998) Glycoconjugate journal. 15:1115-1123.
		755850|MGAT4B	1.2E-169	[Homo sapiens][Transferase] UDP-N-acetylglucosamine:alpha1,3-D-mannoside beta1,4-N-acetylglucosaminyltransferase. Yoshida, A. et al. supra
18	7510368CD1	g837328	5.0E-35	[Homo sapiens] protoporphyrinogen oxidase. Dailey, T. A. et al. (1994) J. Biol. Chem. 26: 813-815.

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
		337132|PPOX	4.1E-36	[Homo sapiens][Oxidoreductase][Cytoplasmic; Mitochondrial] Protoporphyrinogen oxidase, catalyzes the oxidation of protoporphyrinogen IX to protoporphyrin IX in heme biosynthesis, inhibited by the tetrahydrophthalimide and diphenyl ether herbicides; genetic mutation is detected in patients with variegate porphyria. Dailey, T. A. et al. (1995) Arch Biochem Biophys 324:379-384.
		582571|Ppox	1.9E-31	[Mus musculus][Oxidoreductase] Protoporphyrinogen oxidase, catalyzes the oxidation of protoporphyrinogen IX to protoporphyrin IX, inhibited by agents used in light treatment of tumors and by diphenyl ether herbicides; deficiency of human PPOX is associated with variegate porphyria. Camadro, J. M. et al. (1991) Biochem J 277:17-21.
19	7510369CD1	g837328	8.2E-193	[Homo sapiens] protoporphyrinogen oxidase. Dailey, T. A. et al. (1994) J. Biol. Chem. 269:813-815.
		337132|PPOX	6.7E-194	[Homo sapiens][Oxidoreductase][Cytoplasmic; Mitochondrial] Protoporphyrinogen oxidase, catalyzes the oxidation of protoporphyrinogen IX to protoporphyrin IX in heme biosynthesis, inhibited by the tetrahydrophthalimide and diphenyl ether herbicides; genetic mutation is detected in patients with variegate porphyria. Dailey, T. A. et al. supra
		582571|Ppox	2.7E-174	[Mus musculus][Oxidoreductase] Protoporphyrinogen oxidase, catalyzes the oxidation of protoporphyrinogen IX to protoporphyrin IX, inhibited by agents used in light treatment of tumors and by diphenyl ether herbicides; deficiency of human PPOX is associated with variegate porphyria. Taketani, S. et al. (1995) Eur J Biochem 230:760-765.
20	7510377CD1	g6425040	4.3E-164	[Homo sapiens] N-acetylglucosamine-1-phosphodiester alpha-N-acetylglucosaminidase. Kornfeld, R. et al. (1999) J. Biol. Chem. 274:32778-32785.

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
		475837|LOC51172	3.5E-165	[Homo sapiens][Hydrolase] N-acetylglucosamine-1-phosphodiester alpha-N-acetylglucosaminidase (mannose 6-phosphate uncovering enzyme), catalyzes the second step in the synthesis of the mannose 6-phosphate recognition signal on lysosomal enzymes. Kornfeld, R. et al. (1999) J Biol Chem 274:32778-32785.
		419869|Apaa	3.9E-118	[Mus musculus][Hydrolase][Plasma membrane] N-acetylglucosamine-1-phosphodiester alpha-N-acetylglucosaminidase, putative enzyme that functions in the synthesis of the mannose 6-phosphate recognition signal on lysosomal enzymes. Kornfeld, R. et al. supra
21	7510026CD1	g762826	0.0	[Homo sapiens] phospholipase C beta 4. Alvarez, R. A. et al. (1995) Genomics 29:53-61.
		337014|PLCB4	0.0	[Homo sapiens][Hydrolase] Phospholipase C beta 4, member of a G-protein-regulated family of phospholipases that hydrolyze phosphatidylinositol 4,5-bisphosphate to the second messengers inositol 1,4,5-trisphosphate and diacylglycerol; mouse Plcb4 mutants develop ataxia. Alvarez, R. A. et al. supra
		609040|Plcb1	4.9E-197	[Mus musculus][GTPase activating protein; Activator; Hydrolase][Cytoplasmic] Phospholipase C-beta, activated by G-protein coupled signaling receptors and involved in nervous system development; mouse Plcb1 mutants develop epilepsy. Watanabe, M. et al. (1998) Eur J Neurosci 10:2016-2025.
22	7509168CD1	g951352	4.5E-11	[Homo sapiens] glutathione S-transferase A3. Suzuki, T. et al. (1993) Genomics 18:680-686.

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
		625951|GSTA3	3.6E-12	[Homo sapiens][Transferase] Glutathione S-transferase alpha 3, conjugates reduced glutathione to hydrophobic compounds with an electrophilic center, possesses isomerase activity, homodimer catalyzes double-bond isomerization reactions in steroid hormone biosynthesis. Board, P. G. (1998) Biochem J 330:827-831.
		584921|Gsta3	6.8E-11	[Mus musculus][Transferase] Glutathione S-transferase alpha 3, conjugates reduced glutathione to hydrophobic electrophiles, important in xenobiotic metabolism and oxidative stress response, confers resistance to aflatoxin. Buetler, T. M. et al. (1992) Cancer Res 52:314-318.
23	7500607CD1	g2257472	6.8E-41	[Homo sapiens] muscle type carnitine palmitoyltransferase I. Yamazaki, N. et al. (1997) FEBS Lett. 409:401-406.
24	7506079CD1	g632808	1.1E-187	[Homo sapiens] pyruvate carboxylase; pyruvate:carbon dioxide ligase. MacKay, N. et al. (1994) Biochem. Biophys. Res. Commun. 202:1009-1014.
25	7509259CD1	g13195755	4.8E-93	[Homo sapiens] type I 5' iodothyronine deiodinase; 5' DI. Mandel, S. J. et al. (1992) J. Clin. Endocrinol. Metab. 75:1133-1139.
		339284|DIO1	6.3E-96	[Homo sapiens][Hydrolase; Small molecule-binding protein] Iodothyronine deiodinase type I (5' iodothyronine deiodinase), catalyzes the deiodination of thyroxine (T4) to the active 3,3',5-triiodothyronine (T3), and also catalyzes the deiodination of T3 to the inactive T2 (3,3'-diiodothyronine) form. Mandel, S. J. et al. supra
		624452|Dio1	6.0E-79	[Rattus norvegicus][Hydrolase][Plasma membrane] Iodothyronine deiodinase type I (5' iodothyronine deiodinase), catalyzes the deiodination of thyroxine (T4) to the active 3,3',5-triiodothyronine (T3), and also catalyzes the deiodination of T3 to the inactive T2 (3,3'-diiodothyronine) form. Berry, M. J. et al. (1991) Nature 349:438-440.

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
26	7509263CD1	g179795	2.0E-40	[Homo sapiens] carbonic anhydrase II. Forsman, C. et al. (1988) Acta Chem. Scand., B, Org. Chem. Biochem. 42:314-318.
		339166|CA2	1.6E-41	[Homo sapiens][Lyase][Nuclear; Cytoplasmic] Carbonic anhydrase 2, catalyzes the hydration of carbon dioxide to form bicarbonate ion and a proton; mutation of the corresponding gene is associated with osteopetrosis, renal tubular acidosis, and cerebral calcification. Venta, P. J. et al. (1991) Am J Hum Genet 49:1082-1090.
27	7509360CD1	g219974	4.6E-112	[Homo sapiens] cholesterol desmolase cytochrome P-450 (SCC). Morohashi, K. et al. (1987) J. Biochem. 101:879-887.
		334930|CYP11A	5.5E-112	[Homo sapiens][Oxidoreductase][Cytoplasmic; Mitochondrial] Cytochrome P450 enzyme (cholesterol side-chain cleavage), catalyzes the initial and rate limiting step of steroidogenesis, variants may be associated with polycystic ovary syndrome. Sparkes, R. S. et al. (1991) DNA Cell Biol 10:359-365.
		590995|Cyp11a	8.4E-84	[Rattus norvegicus][Oxidoreductase; Transporter; Small molecule-binding protein] Cytochrome P450 enzyme (cholesterol side-chain cleavage), catalyzes the initial and rate limiting step of steroidogenesis; variants of human CYP11A may be associated with polycystic ovary syndrome. Furukawa, A. et al. (1998) J Neurochem 71:2231-2238.
28	7509394CD1	g11414998	1.6E-86	[Homo sapiens] NADPH-cytochrome P-450 reductase. Murakami, H. O. et al. Published Only in DataBase (2000) In press.
		347730|POR	4.4E-87	[Homo sapiens][Oxidoreductase; Transporter; Small molecule-binding protein] Cytochrome P450 oxidoreductase, NADPH-dependent and heme-binding monooxygenase that metabolizes steroids, fatty acids, and xenobiotics, regulates cytochrome P450 activity and protein levels. Shephard, E. A. et al. (1989) Ann Hum Genet 291-301.

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
		711444|Por	2.6E-80	[Rattus norvegicus][Oxidoreductase][Endoplasmic reticulum; Cytoplasmic] Cytochrome P450 oxidoreductase, NADPH-dependent and heme-binding monooxygenase that may metabolize steroids, fatty acids and xenobiotics; upregulated by thyroid hormone in liver tissue. Sutter, T. R. et al. (1990) J Biol Chem 265:16428-16436.
29	7581076CD1	g15680046	1.2E-31	[Homo sapiens] Similar to lactate dehydrogenase 1, A chain.
		345096|LDHB	5.5E-32	[Homo sapiens][Oxidoreductase] Lactate dehydrogenase heart subunit, catalyzes the reversible NAD-dependent interconversion of pyruvate to L-lactate; mutations in the corresponding gene are associated with lactate dehydrogenase B deficiency. Sakai, I. et al. (1987) Biochem J 248:933-936.
		590313|Ldhb	1.2E-31	[Rattus norvegicus][Oxidoreductase] Lactate dehydrogenase heart subunit, catalyzes the reversible NAD-dependent interconversion of pyruvate to L-lactate; mutations in the corresponding human LDHB gene are associated with lactate dehydrogenase B deficiency. Tsuji, S. et al. (1994) Proc Natl Acad Sci U S A 91:9392-9396.
30	7504551CD1	340938|NDUFS2	1.3E-11	[Homo sapiens] [Oxidoreductase; Small molecule-binding protein] [Cytoplasmic; Mitochondrial; Mitochondrial inner membrane] Subunit of the NADH-ubiquinone oxidoreductase (complex I), a multiprotein complex that transports electrons from NADH to ubiquinone, present within the complex I iron-sulfur protein (IP) fraction, has a predicted PKC phosphorylation site Loeffen, J. et al. Biochem Biophys Res Commun 247, 751-8 (1998).
31	7500652CD1	g3746533	1.2E-49	[Homo sapiens] ACAT related gene product 1 Oelkers, P. et al. J. Biol. Chem. 273, 26765-26771 (1998)

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
	7500652CD1	432898|DGAT1	9.7E-51	[Homo sapiens] [Transferase] Acyl CoA:diacylglycerol acyltransferase, an enzyme involved in the biosynthesis of triacylglycerol, may play a role in fat body development and lactation Cases, S. et al. Proc Natl Acad Sci U S A 95, 13018-23 (1998) Oelkers, P. et al. J Biol Chem 273, 26765-71 (1998)
	7500652CD1	584645|Dgat1	7.4E-37	[Mus musculus] [Transferase] Acyl CoA:diacylglycerol acyltransferase, catalyzes the transfer of an acyl group from acyl-CoA to diacylglycerol forming triacylglycerol, plays a role in lactation, fat absorption, adipose tissue formation, and lipoprotein assembly Smith, S. J. et al. Nat Genet 25, 87-90 (2000) Coburn, C. T. et al. J Biol Chem 275, 32523-9 (2000 Tansey, J. T. et al. Proc Natl Acad Sci U S A 98, 6494-9. (2001)
32	7500900CD1	g1082036	7.6E-215	[Homo sapiens] fatty aldehyde dehydrogenase De Laurenzi, V. et al. Nature Genet 12, 52-57 (1996)
	7500900CD1	339660|ALDH3A2	6.1E-216	[Homo sapiens] [Oxidoreductase] Aldehyde dehyrogenase 3 family member A2, oxidizes long chain fatty aldehydes and leukotrienes; mutation of the corresponding gene causes Sjogren Larsson syndrome, a disorder marked by mental retardation, spasticity, and ichthyosis De Laurenzi, V. et al. Nat Genet 12, 52-7 (1996) Chang, C. et al.Genomics 40, 80-5 (1997)
	7500900CD1	329142|Rn.9113	3.3E-192	[Rattus norvegicus] [Oxidoreductase] [Endoplasmic reticulum; Cytoplasmic; Plasma membrane] Microsomal aldehyde dehydrogenase class 3, oxidizes long and medium chain aliphatic aldehydes derived from lipid metabolism; mutation of the corresponding human ALDH3A2 gene causes Sogren Larsson syndrome Miyauchi, K. et al. J Biol Chem 266, 19536-42 (1991) Lindahl, R. et al. Biochem Pharmacol 41, 1583-7. (1991)

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
33	7501398CD1	475555|LOC51005	1.3E-48	[Homo sapiens] Member of the N-acetylglucosamine-6-phosphate deacetylase family, has high similarity to uncharacterized C. elegans F59B2.3
34	7501417CD1	g18073364	2.4E-35	[Homo sapiens] (AJ277883) sialidase
	7501417CD1	477402|Neu3	5.4E-11	[Mus musculus] [Lyase; Hydrolase] Neuraminidase 3 (ganglioside sialidase), catalyzes the removal of sialic acid from gangliosides in ganglioside catabolism and neurogenesis; human NEU3 deficiency is associated with the lysosomal disorder sialidosis Hasegawa, T. et al. J Biol Chem 275, 8007-15 (2000)
35	7501472CD1	g6649583	1.1E-240	[Homo sapiens] kidney and liver proline oxidase 1
	7501472CD1	618062|HSPOX1	8.5E-242	[Homo sapiens] [Oxidoreductase] Member of the proline dehydrogenase family, has a region of strong similarity to a region of NPHS1 nephrin, which is expressed in renal glomeruli and for which mutation of corresponding gene causes congenital nephrotic syndrome
	7501472CD1	704217|Pox1	7.9E-136	[Mus musculus] [Oxidoreductase] Member of the proline dehydrogenase family
36	7501489CD1	g1747521	1.5E-21	[Homo sapiens] microsomal glutathione S-transferase 2 Jakobsson, P. J. et al. J. Biol. Chem. 271, 22203-22210 (1996)
	7501489CD1	336410|MGST2	1.2E-22	[Homo sapiens] [Transferase] [Endoplasmic reticulum; Cytoplasmic; Microsomal fraction; Unspecified membrane] Microsomal glutathione S-transferase 2, catalyzes the production of LTC4 from LTA4 and reduced glutathione, plays a role in signal transduction Jakobsson, P. J. et al. J Biol Chem 271, 22203-10 (1996) Scoggan, K. A. et al. J Biol Chem 272, 10182-7 (1997)
37	7501555CD1	g12653409	1.2E-106	[Homo sapiens] quinone oxidoreductase homolog

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
	7501555CD1	341052|PIG3	9.5E-108	[Homo sapiens] [Oxidoreductase] Protein with low similarity to Zeta crystallins, which are lens crystallins with NADPH: quinone oxidoreductase activity, member of the zinc-containing alcohol dehydrogenase family, which oxidize ethanol to acetaldehyde with concomitant reduction of NAD
	7501555CD1	370850| SPCC285.01c	2.2E-21	[Schizosaccharomyces pombe] Quinone oxidoreductase
38	7501561CD1	g9739150	1.4E-34	[Homo sapiens] acyl-CoA dehydrogenase 8
	7501561CD1	568940|ACAD8	1.1E-35	[Homo sapiens] [Oxidoreductase] [Cytoplasmic; Mitochondrial] Member of the acyl-Coenzyme A dehydrogenase family that alpha,beta-dehydrogenates acyl-CoA esters and transfers electrons to flavoproteins Telford, E. A. et al. Biochim Biophys Acta 1446, 371-6 (1999).
39	7506108CD1	g12655177	2.8E-84	[Homo sapiens] CDP-diacylglycerol--inositol 3-phosphatidyltransferase (phosphatidylinositol synthase)
	7506108CD1	343656|CDIPT	2.2E-85	[Homo sapiens] [Transferase; Small molecule-binding protein] Phosphatidylinositol synthase, has both phosphatidylinositol synthase and phosphatidylinositol:inositol exchange activity Lykidis, A. et al. J Biol Chem 272, 33402-9 (1997) Piatti, E. et al. Biochimie 81, 1011-4. (1999)
	7506108CD1	330796|Rn.10598	8.6E-82	[Rattus norvegicus] [Transferase] Phosphatidylinositol synthase (CDP-1,2-diacyl-sn-glycerol: 3-phosphatidyltransferase), converts CDP-diacylglycerol and myo-inositol to phosphatidylinositol and CMP Tanaka, S. et al. FEBS Lett 393, 89-92 (1996)
				Claro, E. et al. J Neurochem 58, 2155-61 (1992)
40	7506123CD1	g12654301	1.3E-92	[Homo sapiens] Similar to phosphoserine aminotransferase
	7506123CD1	760403|PSA	1.0E-93	[Homo sapiens] [Transferase] Member of the aminotransferases class-V family of pyridoxal-phosphate-dependent enzymes, has moderate similarity to S. cerevisiae Ser1p, which is a 3-phosphoserine transaminase involved in synthesis of serine from 3-phosphoglycerate

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
	7506123CD1	243845|F26H9.5	1.6E-50	[Caenorhabditis elegans] [Transferase] Putative phosphoserine aminotransferase
41	7506248CD1	g13623199	0.0	[Homo sapiens] ATP citrate lyase
	7506248CD1	590837|Acly	0.0	[Rattus norvegicus] [Lyase] [Cytoplasmic] ATP citrate lyase, catalyzes the formation of acetyl-CoA and oxaloacetate from citrate and CoA with the concomitant hydrolysis of ATP to ADP and phosphate, involved in cholesterol and fatty acid synthesis, nutritionally regulated Elshourbagy, N. A. et al. J Biol Chem 265, 1430-5 (1990) Moon, Y. A. et al. J Biol Chem 275, 30280-6 (2000)
	7506248CD1	334006|ACLY	0.0	[Homo sapiens] [Lyase] ATP citrate lyase, catalyzes the formation of acetyl-CoA and oxaloacetate from citrate and CoA with the concomitant hydrolysis of ATP to ADP and phosphate, involved in cholesterol and fatty acid synthesis, nutritionally regulated Elshourbagy, N. A. et al. Eur J Biochem 204, 491-9 (1992) Sato, R. et al. J Biol Chem 275, 12497-502 (2000)
42	7506347CD1	g182671	5.8E-255	[Homo sapiens] flavin-containing monooxygenase Dolphin, C. et al. J. Biol. Chem. 266, 12379-12385 (1991)
	7506347CD1	341488|FMO1	4.6E-256	[Homo sapiens] [Oxidoreductase] Flavin containing monooxygenase, member of a family that catalyzes the monooxygenation of xenobiotic soft nucleophiles Dolphin, C. et al. J Biol Chem 266, 12379-85 (1991) Cashman, J. R. Curr Drug Metab 1, 181-91. (2000)
	7506347CD1	584811|Fmo1	3.4E-220	[Mus musculus] [Oxidoreductase] Flavin-containing monooxygenase, member of a family that catalyzes the monooxygenation of xenobiotic soft nucleophiles, active toward methimazole and phorate Cherrington, N. J. et al. J Biochem Mol Toxicol 12, 205-12 (1998)
43	7509172CD1	g181326	9.6E-57	[Homo sapiens] cytochrome P-450 1 Okino, S. T. et al. J. Biol. Chem. 262, 16072-16079 (1987)

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
	7509172CD1	703979|CYP2C8	7.7E-58	[Homo sapiens] [Oxidoreductase; Transporter; Small molecule-binding protein] [Endoplasmic reticulum; Cytoplasmic] Member of the cytochrome P450 IIC subfamily in the heme-binding monooxygenase superfamily that metabolizes steroids, fatty acids, and xenobiotics Furuya, H. et al. Mol Pharmacol 40, 375-82 (1991) Kimura, S. et al. Nucleic Acids Res 15, 10053-4 (1987)
	7509172CD1	584613|Cyp2c37	1.1E-40	[Mus musculus] [Oxidoreductase; Transporter; Small molecule-binding protein] Cytochrome P450 2C37, metabolizes arachidonic acid to 12-hydroxyeicosatetraenoic acid, expressed in liver Luo, G. et al. Arch Biochem Biophys 357, 45-57 (1998) Tsao, C. C. et al. Mol Pharmacol 58, 279-87 (2000)
44	7510421CD1	g15928817	1.3E-18	[Homo sapiens] Similar to pyrroline 5-carboxylate reductase isoform
	7510421CD1	568062|PYCR1	4.3E-16	[Homo sapiens] [Oxidoreductase] Pyrroline-5-carboxylate reductase 1, catalyzes the NAD(P)H dependent reduction of pyrroline-5-carboxylate to proline in the final step of proline synthesis, may also function NADP+ production in erythrocytes Dougherty, K. M. et al. J Biol Chem 267, 871-5 (1992) Merrill, M. J. et al. J Biol Chem 264, 9352-8 (1989)
45	7504625CD1	g6318546	2.1E-45	[Homo sapiens] retinal short-chain dehydrogenase/reductase retSDR3 Haeseleer, F. et al. Meth. Enzymol. 316 (1999) In press
	7504625CD1	475835|LOC51171	1.7E-46	[Homo sapiens] Retinal short-chain dehydrogenase/reductase
	7504625CD1	462738|arp2	2.9E-14	[Aspergillus fumigatus] [Oxidoreductase] 1,3,6,8-tetrahydroxynaphthalene reductase, involved in conidial pigment biosynthesis
46	7504776CD1	g386992	4.0E-188	[Homo sapiens] cytochrome P450c17 Picado-Leonard, J. et al. DNA 6, 439-448 (1987)

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
	7504776CD1	334932|CYP17	2.3E-190	[Homo sapiens] [Oxidoreductase; Transporter; Small molecule-binding protein] [Endoplasmic reticulum; Microsomal fraction; Cytoplasmic] Steroid 17 alpha-hydroxylase/C17-20 lyase, a cytochrome P450 enzyme that catalyzes both 17-hydroxylase and 17,20-lyase activities in steroid hormone biosynthesis, an autoantigen in adrenocortical failure Krohn, K. et al. Lancet 339, 770-3 (1992) Geller, D. H. et al. Nat Genet 17, 201-5 (1997)
	7504776CD1	590011|Cyp17	4.5E-144	[Rattus norvegicus] [Oxidoreductase] Steroid 17 alpha-hydroxylase/C17-20 lyase, a cytochrome P450 enzyme that acts in corticosteroid and androgen production, converts 17 alpha-hydroxypregnenolone and 17 alpha-hydroxyprogesterone into dehydroepiandrosterone and androstenedione Nishihara, M. et al. Biochem Biophys Res Commun 154, 151-8 (1988) Fevold, H. R. et al. Mol Endocrinol 3, 968-75 (1989)
47	7504927CD1	g9802312	2.9E-76	[Homo sapiens] NADH-ubiquinone dehydrogenase 1 beta subcomplex Ye, Z. et al. Biochem. Biophys. Res. Commun. 275, 223-227 (2000)
	7504927CD1	432608|NDUFB9	2.4E-77	[Homo sapiens] [Oxidoreductase] [Cytoplasmic; Unspecified membrane; Mitochondrial; Mitochondrial inner membrane] NADH dehydrogenase ubiquinone 1 beta subcomplex 9 (22 kDa), a subunit of the NADH-ubiquinone oxidoreductase (complex I), a multiprotein complex that transports electrons from NADH to ubiquinone Gu, J. Z. et al. Genomics 35, 6-10 (1996).
48	7505010CD1	g16307450	1.2E-199	[Homo sapiens] multifunctional polypeptide similar to SAICAR synthetase and AIR carboxylase
	7505010CD1	343344|ADE2H1	9.9E-201	[Homo sapiens] [Lyase; Ligase] SAICAR (phosphoribosylaminoimidazole-succinocarboxamide) synthetase and AIR (phosphoribosylaminoimidazole) carboxylase, a bifunctional protein required for de novo purine biosynthesis Minet, M. et al. Curr Genet 18, 287-91 (1990) Chassin, D. et al. Cancer Res 54, 5217-23 (1994)

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
	7505010CD1	773443|Paics	2.1E-196	[Rattus norvegicus] [Lyase; Ligase] SAICAR (phosphoribosylaminoimidazole-succinocarboxamide) synthetase and AIR (phosphoribosylaminoimidazole) carboxylase, a bifunctional protein required for de novo purine biosynthesis Iwahana, H. et al. Biochim Biophys Acta 1261, 369-80 (1995)
49	7505173CD1	g4567042	2.1E-53	[Homo sapiens] NADH:ubiquinone oxidoreductase SDAP subunit Loeffen, J. L. C. M. et al. Biochem. Biophys. Res. Commun. 253, 415-422 (1998)
	7505173CD1	341720|NDUFAB1	1.7E-54	[Homo sapiens] [Oxidoreductase] [Cytoplasmic; Mitochondrial; Mitochondrial inner membrane] NADH-ubiquinone oxidoreductase subunit of alpha-beta subcomplex 1 (8 kD), a probable acyl carrier component of the multiprotein complex that transports electrons from NADH to ubiquinone, present within the hydrophobic fraction of complex I Runswick, M. J. et al. FEBS Lett 286, 121-4 (1991) Loeffen, J. L. et al. Biochem Biophys Res Commun 253, 415-22 (1998)
	7505173CD1	443533|Y56A3A.O	9.0E-18	[Caenorhabditis elegans] [Oxidoreductase] [Mitochondrial] Putative NADH-ubiquinone oxidoreductase acyl-carrier subunit, protein with strong similarity to C. elegans F16B4.6 and Drosophila ND-AcC protein
50	7510061CD1	g2792518	0.0	[Homo sapiens] heparan glucosaminyl N-deacetylase/N-sulfotransferase-2
	7510061CD1	336568|NDST2	0.0	[Homo sapiens] [Transferase; Hydrolase] Heparan glucosaminyl N-deacetylase/N-sulphotransferase-2, has a role in heparin biosynthesis Humphries, D. E. et al. Biochem J 332, 303-7 (1998) Aikawa, J. et al. J Biol Chem 274, 2690-5 (1999)
	7510061CD1	586593|Ndst2	0.0	[Mus musculus] [Transferase; Hydrolase] [Golgi; Endoplasmic reticulum; Cytoplasmic] Heparin glucosaminyl N-deacetylase/N-sulphotransferase-2, has a role in heparin biosynthesis Aikawa, J. et al. J Biol Chem 276, 5876-82. (2001) Eriksson, I. et al. J Biol Chem 269, 10438-43 (1994)
51	7510091CD1	g12314172	7.9E-62	[Homo sapiens] bA13B9.1 (novel protein similar to a glycosyltransferase)

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
	7510091CD1	743480|FLJ14511	6.5E-63	[Homo sapiens] Member of the glycosyl transferases group 1 family, which transfer nucleotide-linked sugars to acceptor substrates, has strong similarity to uncharacterized mouse 1300013N08Rik
	7510091CD1	703080|pi010	4.4E-21	[Schizosaccharomyces pombe] [Transferase] Member of the glycosyl transferases group 1 family, which transfer nucleotide-linked sugars to acceptor substrates, has moderate similarity to S. cerevisiae Alg2p, which is a mannosyltransferase involved in N-glycosylation Machida, M. et al. Yeast 16, 71-80 (2000)
52	7510109CD1	g12846107	6.6E-248	[Mus musculus] (AK010568) data source:SPTR, source key:Q9RYW0, evidence:ISS~putative~related to ACYL-COA DEHYDROGENASE, PUTATIVE Carninci, P. et al. Meth. Enzymol. 303, 19-44 (1999) Carninci, P. et al. Genome Res. 10, 1617-1630 (2000) Shibata, K. et al. Genome Res. 10, 1757-1771 (2000)
	7510109CD1	692048|MGC5601	1.7E-183	[Homo sapiens] Protein with low similarity to short/branched chain acyl-Coenzyme A dehydrogenase (human ACADSB), which oxidizes branched chain and straight chain acyl-CoAs in the metabolism of fatty acids or branched chain amino acids
	7510109CD1	313841|K09H11.1	7.4E-181	[Caenorhabditis elegans] [Oxidoreductase] Protein containing a putative acyl-CoA dehydrogenase domain, has strong similarity at the C-terminus to human ACADSB, an acyl-CoA dehydrogenase
53	7510121CD1	g9622124	4.5E-114	[Homo sapiens] androgen-regulated short-chain dehydrogenase/reductase 1 Lin, B. et al. Cancer Res. 61, 1611-1618 (2001)
	7510121CD1	475723|LOC51109	3.7E-115	[Homo sapiens] Protein containing EGF-like domains, which are found in some secreted proteins and extracellular domains of transmembrane proteins
	7510121CD1	626578|Mdt1	2.8E-101	[Mus musculus] cell line MC/9.IL4 derived transcript 1

Table 2

Polypeptide SEQ ID NO:	Incyte Polypeptide ID	GenBank ID NO: or PROTEOME ID NO:	Probability Score	Annotation
54	7510797CD1	g3818572	0.0	[Homo sapiens] cAMP-specific phosphodiesterase 8B; PDE8B1; 3',5'-cyclic nucleotide phosphodiesterase Hayashi, M. et al. Biochem. Biophys. Res. Commun. 250, 751-756 (1998)
	7510797CD1	347910|PDE8B	0.0	[Homo sapiens] [Hydrolase] cAMP-specific phosphodiesterase, highly expressed in the thyroid Hayashi, M. et al. Biochem Biophys Res Commun 250, 751-6 (1998).
	7510797CD1	582401|Pde8a	3.0E-263	[Mus musculus] [Hydrolase] cAMP-specific phosphodiesterase, highly expressed in testis Soderling, S. H. et al. Proc Natl Acad Sci U S A 95, 8991-6 (1998)
55	7504944CD1	g15559560	0.0	[Homo sapiens] Similar to glucuronidase, beta
	7504944CD1	335706|GUSB	0.0	[Homo sapiens] [Hydrolase] [Lysosome/vacuole; Cytoplasmic; Extracellular (excluding cell wall)] Beta-glucuronidase (beta-D-glucuronoside glucuronosohydrolase), a lysosomal hydrolase that degrades glucuronic acid-containing glycosaminoglycans; deficiency causes mucopolysaccharidosis VII (Sly syndrome) Oshima, A. et al. Proc Natl Acad Sci U S A 84, 685-9 (1987) Moullier, P. et al. Nat Genet 4, 154-9 (1993)
	7504944CD1	591117|Gusb	2.6E-257	[Rattus norvegicus] [Hydrolase] [Lysosome/vacuole; Endoplasmic reticulum; Cytoplasmic] Beta-glucuronidase (beta-D-glucuronoside glucuronosohydrolase), a lysosomal hydrolase that degrades glucuronic acid-containing glycosaminoglycans; deficiency of human GUSB causes mucopolysaccharidosis VII (Sly syndrome) Nishimura, Y. et al. Proc Natl Acad Sci U S A 83, 7292-6 (1986) Shipley, J. M. et al. J Biol Chem 268, 12193-8. (1993)

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
1	7509350CD1	355	S64 S154 S295 S304 T141 T196 T219 T319 Y220		signal_cleavage: M1-P18	SPSCAN
					Signal Peptide: M1-P18	HMMER
					Cytochrome P450: P44-Q334	HMMER_PFAM
					Cytosolic domain: D33-P355 Transmembrane domain: A10-A32 Non-cytosolic domain: M1-A9	TMHMMER
					Mitochondrial P450 signature PR00408: S129-R139, N185-Y203, A311-A328	BLIMPS_PRINTS
					E-class P450 group I signature PR00463: V72-I91, L96-R117, A187-D205, N300-S317, L320-A346	BLIMPS_PRINTS
					CYTOCHROME P450 MONOOXYGENASE OXIDOREDUCTASE HEME ELECTRON TRANSPORT MEMBRANE MICROSOME ENDOPLASMIC PD000021: I184-Q334, K152-M285	BLAST_PRODOM
					CYTOCHROME P450 DM00022 |P52786|83-492: S87-Q334 |A46588|15-478: L30-P349 |I49625|39-475: L53-P349 |P24460|49-475: Q63-V352	BLAST_DOMO
2	7509325CD1	139	S23 S95 S103 S131	N56	signal_cleavage: M1-R21	SPSCAN
					Signal Peptide: M1-S18, M1-R21, M1-Q22, M1-R25	HMMER
					Cytochrome P450: P30-G111	HMMER_PFAM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					Cytosolic domain: Q22-A139 Transmembrane domain: F4-R21 Non-cytosolic domain: M1-P3	TMHMMER
					E-class P450 group I signature PR00463: S58-F77, A82-S103	BLIMPS_PRINTS
					CYTOCHROME P450 MONOOXYGENASE OXIDOREDUCTASE HEME ELECTRON TRANSPORT MEMBRANE MICROSOME ENDOPLASMIC PD000021: P30-P101	BLAST_PRODOM
					CYTOCHROME P450 DM00022 |I48164|11-474: L11-G111 |P10632|48-474: K48-G111 |A46588|15-478: V8-G111 |P33262|48-474: K48-G111	BLAST_DOMO
3	7509337CD1	259	S92 S135 S210 T118 T182 T191 T202 T224	N28 N203	Sulfotransferase protein: E17-K247	HMMER_PFAM
					Sulfotransferase protein PF00685: F35-D67, P116-K161, F219-M248, V171-S208	BLIMPS_PFAM
					TRANSFERASE SULFOTRANSFERASE STEROID METABOLISM HYDROXYSTEROID ALCOHOL PHENOL ESTROGEN PROTEIN STEROID-BINDING PD001218: D27-P88, A93-K247	BLAST_PRODOM
					PAPS BINDING SITE BINDING DM00981 |A55451|5-296: L13-D98, A93-L259 |P52847|5-298: L13-E84, A93-L259 |P52840|1-291: L13-P91, A93-F255 |A44011|6-297: E15-E84, R94-L259	BLAST_DOMO

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
4	7509353CD1	450	S210 S216 S294 S344 T141 Y255	N414	PUTATIVE HYDROXYPYRUVATE REDUCTASE PLASMID OXIDOREDUCTASE NADP PROTEIN GLYCERATE KINASE 440AA PD014236: P113-L447	BLAST_PRODOM
5	7509354CD1	387	S243 S266 S323 S336 T52 T112 T227 Y16	N2	Pyridoxal-dependent decarboxylase conse: V147-K321, P35-Q145	HMMER_PFAM
					Aromatic-L-amino-acid decarboxylase signature PR00800: F6-E25, V29-A46, P47-P66, S73-D92, L94-T112, E113-F132, G140-N160, P259-R274, D302-K321	BLIMPS_PRINTS
					DECARBOXYLASE LYASE PYRIDOXAL PHOSPHATE MULTIGENE FAMILY DOPA GLUTAMATE ACID AROMATIC L-AMINO ACID PD001960: P35-K321	BLAST_PRODOM
					PYRIDOXAL PHOSPHATE DECARBOXYLASE PROTEIN LYASE AMINOTRANSFERASE BIOSYNTHESIS TRANSFERASE GLYCINE DEHYDROGENASE PD000897: E139-S247	BLAST_PRODOM
					DECARBOXYLASE LYASE PYRIDOXAL PHOSPHATE DOPA AROMATIC L-AMINO ACID DDC CATECHOLAMINE METABOLISM PROTEIN PD011223: S323-L382	BLAST_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					DDC / GAD / HDC / TYRDC PYRIDOXAL-PHOSPHATE ATTACHMENT SITE DM00568 |P27718|1-477: M1-T152, M146-E385 |P05032|28-501: M1-T152, L128-R386 |A25697|36-510: M1-T152, L128-R386 |P48861|1-475: M1-T152, M146-K376	BLAST_DOMO
					DDC / GAD / HDC / TyrDC pyridoxal-phosphate attachment site: S203-K224	MOTIFS
6	7509385CD1	442	S109 S155 S298 S321 S378 S391 T52 T74 T282 Y16	N2	Pyridoxal-dependent decarboxylase conserved domain: A68-K376, P35-G67	HMMER_PFAM
					DDC / GAD / HDC / TyrDC pyridoxal-phosphate attachment site proteins BL00392: W229-G238	BLIMPS_BLOCKS
					Aromatic-L-amino-acid decarboxylase signature PR00800: I56-T74, E75-F94, G102-R122, P314-R329, D357-K376	BLIMPS_PRINTS
					DECARBOXYLASE LYASE PYRIDOXAL PHOSPHATE MULTIGENE FAMILY DOPA GLUTAMATE ACID AROMATIC L-AMINO ACID PD001960: P35-G67, A69-K376	BLAST_PRODOM
					PYRIDOXAL PHOSPHATE DECARBOXYLASE PROTEIN LYASE AMINOTRANSFERASE BIOSYNTHESIS TRANSFERASE GLYCINE DEHYDROGENASE PD000897: L145-S302	BLAST_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					DECARBOXYLASE LYASE PYRIDOXAL PHOSPHATE DOPA AROMATIC L-AMINO ACID DDC CATECHOLAMINE METABOLISM PROTEIN PD011223: S378-L437	BLAST_PRODOM
					DDC / GAD / HDC / TYRDC PYRIDOXAL-PHOSPHATE ATTACHMENT SITE DM00568 |P27718|1-477: M1-E440 |P05032|28-501: M1-G67, A69-R441 |P48861|1-475: M1-G67, A69-E431 |A25697|36-510: M1-G67, A69-R441	BLAST_DOMO
					DDC / GAD / HDC / TyrDC pyridoxal-phosphate attachment site: S258-K279	MOTIFS
7	7509216CD1	161	S58 T5 T14 T64 T86 T143		signal_cleavage: M1-G51	SPSCAN
					Signal Peptide: M32-P47, M32-Q50, M32-G51, M32-L54, M25-G51, M32-A48	HMMER
					Phospholipase A2: G52-W93	HMMER_PFAM
					Phospholipase A2: G52-S152	HMMER_SMART
					Phospholipase A2 histidine proteins BL00118: G52-T64, Y75-E102	BLIMPS_BLOCKS
					Phospholipase A2 active sites signatures: Y75-L96	PROFILESCAN
					Phospholipase A2 signature PR00389: L53-V63, A68-T86	BLIMPS_PRINTS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					PHOSPHOLIPASE A2 ASPARTIC ACID DM00093 |P39877|20-134: G51-W93 |P51433|20-134: G51-L113 |P49121|17-128: L53-D92 |P14421|1-112: L53-D92	BLAST_DOMO
8	7509376CD1	126	S58 T5 T14 T64 T86		signal_cleavage: M1-G51	SPSCAN
					Signal Peptide: M32-P47	HMMER
					Signal Peptide: M32-Q50, M32-G51, M32-L54, M25-G51, M32-A48	HMMER
					Phospholipase A2: G52-W93	HMMER_PFAM
					Phospholipase A2 histidine proteins BL00118: G52-T64, Y75-E102	BLIMPS_BLOCKS
					Phospholipase A2 active sites signatures: Y75-L96	PROFILESCAN
					Phospholipase A2 signature PR00389: L53-V63, A68-T86, P87-G105	BLIMPS_PRINTS
					PHOSPHOLIPASE A2 ASPARTIC ACID DM00093 |P51433|20-134: G51-L113 |P49121|17-128: L53-D92 |P39877|20-134: G51-W93 |P14421|1-112: L53-D92	BLAST_DOMO
9	7501927CD1	418	S139 S186 S187 S233 S305 S314 S388 T68 T106 T277	N112 N168	signal_cleavage: M1-A32	SPSCAN
					Signal Peptide: M16-A36	HMMER

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					Cytochrome P450: P52-G417	HMMER_PFAM
					Cytosolic domain: T38-G87 Transmembrane domains: A15-W37, F88-I107 Non-cytosolic domains: M1-V14, R108-L418	TMHMMER
					Mitochondrial P450 signature PR00408: S139-R149, E328-A345, K346-Q359, A376-P394	BLIMPS_PRINTS
					E-class P450 group II signature PR00464: G141-K161, L197-Q215, D317-A345, K346-K363, Q377-T397	BLIMPS_PRINTS
					CYTOCHROME P450 OXIDOREDUCTASE MONOOXYGENASE ELECTRON TRANSPORT MEMBRANE HEME MICROSOME ENDOPLASMIC PD008467: M1-L74	BLAST_PRODOM
					CYTOCHROME P450 MONOOXYGENASE OXIDOREDUCTASE HEME ELECTRON TRANSPORT MEMBRANE MICROSOME ENDOPLASMIC PD000021: Q79-L226 I271-A399	BLAST_PRODOM
					C44C10.2 PROTEIN PD035312: L132-L303	BLAST_PRODOM
					CYTOCHROME P450 DM00022 |Q08477|108-511: A113-L418 |P33274|108-511: R108-L418 |JC4532|108-511: R108-L418 |P51870|108-513: I110-G417	BLAST_DOMO
10	7503274CD1	424	S37 S56 S232 S321 S328 T187 T393 Y144 Y174		signal_cleavage: M1-G42	SPSCAN
					Cytochrome P450: P61-K409	HMMER_PFAM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					Mitochondrial P450 signature PR00408: W133-T148, T149-Q159, Y210-C228, A335-S352, K353-V366, A381-V399	BLIMPS_PRINTS
					E-class P450 group II signature PR00464: G151-A171, Y211-L229, E324-S352, K353-P370, H382-T402	BLIMPS_PRINTS
					ELECTRON TRANSPORT OXIDOREDUCTASE STEROL 26HYDROXYLASE MITOCHONDRIAL PRECURSOR VITAMIN D3 25HYDROXYLASE PD013580: M1-L79	BLAST_PRODOM
					CYTOCHROME P450 MONOOXYGENASE OXIDOREDUCTASE HEME ELECTRON TRANSPORT MEMBRANE MICROSOME ENDOPLASMIC PD000021: I60-E223, Q319-S404	BLAST_PRODOM
					CYTOCHROME P450 DM00022 |Q02318|101-521: G101-L411 |P17178|103-524: G101-E410 |I49442|102-506: V106-K409 |Q07973|102-505: V106-K409	BLAST_DOMO
11	7509104CD1	351	S55 S67 S195 S230 S330 T7 T93	N119 N252	Aldehyde dehydrogenase family: M1-S315	HMMER_PFAM
					Aldehyde dehydrogenases glutamic acid proteins BL00687: G18-R35, E103-K144, P284-G294	BLIMPS_BLOCKS
					ALDEHYDE DEHYDROGENASE-LIKE PROTEIN PD065537: V146-A340	BLAST_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					DEHYDROGENASE OXIDOREDUCTASE ALDEHYDE NAD PROTEIN CLASS SEMI-ALDEHYDE PRECURSOR TRANSIT PEPTIDE PD000218: R12-S265, E141-T483	BLAST_PRODOM
					ALDEHYDE DEHYDROGENASE 7 EC 1.2.1.5 OXIDOREDUCTASE NAD PD065535: S315-L351	BLAST_PRODOM
					ALDEHYDE DEHYDROGENASES GLUTAMIC ACID DM00100 |P43353|1-408: M1-S161, V136-G292 |P30838|1-408: M1-Q160, A131-G292 |P12693|1-432: L8-I154, C134-G292 |P30840|29-471: R9--V158, D159-G292	BLAST_DOMO
12	7509996CD1	366	S65 S94 S187 S213 T152 T295	N185 N208 N226	signal_cleavage: M1-P18	SPSCAN
					Signal Peptide: M1-P18, M1-C22, M1-G24	HMMER
					Cytochrome P450: P28-V324	HMMER_PFAM
					E-class P450 group I signature PR00463: Q57-V76, L81-L102, A174-D192, H291-T308, V311-Q337	BLIMPS_PRINTS
					CYTOCHROME P450 MONOOXYGENASE OXIDOREDUCTASE HEME ELECTRON TRANSPORT MEMBRANE MICROSOME ENDOPLASMIC PD000021: L34-V324	BLAST_PRODOM
					CYTOCHROME P450 DM00022 |P05093|69-479: G69-V324 |P05185|69-479: G69-V324 |S52756|69-479: G69-V324 |P11715|69-478: G69-V324	BLAST_DOMO

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
13	7510030CD1	246	S109 S183 S208 S234 S239 T69 T230 T232 Y94 Y192	N193 N228	signal_cleavage: M1-G38	SPSCAN
					Signal Peptide: M1-G38	HMMER
					Chitinases family 18 proteins BL01095: G85-A95, M135-V146	BLIMPS_BLOCKS
					DINACETYLCHITOBIASE PRECURSOR HYDROLASE GLYCOSIDASE SIGNAL LYSOSOME GLYCOPROTEIN PD034983: D40-Y153	BLAST_PRODOM
					Chitinases family 18 active site: M135-E143	MOTIFS
14	7510062CD1	463	S41 S108 S215 S235 S244 S358 S448 S452 T80 T126 T301 T364 T366 T441 Y312	N233 N350 N400	signal_cleavage: M1-G32	SPSCAN
					Cytosolic domain: M1-R18 Transmembrane domain: L19-S41 Non-cytosolic domain: P42-R463	TMHMMER
					SULFATE TRANSFERASE N-DEACETYLASE/N-SULFOTRANSFERASE SULFOTRANSFERASE GLUCOSAMINYL HEPARAN N HSST DEACETYLASE/N-SULFOTRANSFERASE TRANSMEMBRANE PD011877: V87-A431	BLAST_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					HEPARIN SULFATE N-DEACETYLASE/N-SULFOTRANSFERASE N HSST N-HEPARIN SULFOTRANSFERASE GLUCOSAMINYL DEACETYLASE/N-SULFOTRANSFERASE TRANSFERASE PD043426: M1-V86	BLAST_PRODOM
					SIGNAL-ANCHOR TRANSMEM DM07899 |P52850|1-882: M1-A431 |P52848|1-881: R9-A431	BLAST_DOMO
					Cell attachment sequence: R121-D123	MOTIFS
15	7510217CD1	165	S32 T6 T93 T99 Y108	N31 N115	signal_cleavage: M1-G25	SPSCAN
					Signal Peptide: M1-G25, M1-A30	HMMER
					Sulfatases proteins BL00523: L37-C53, C84-R95, G127-F137	BLIMPS_BLOCKS
					Sulfatases signatures: N106-G161	PROFILESCAN
					HYDROLASE ARYLSULFATASE PRECURSOR SIGNAL GLYCOPROTEIN LYSOSOME PROTEIN SULPHOHYDROLASE MUCOPOLYSACCHARIDOSIS SULFATASE PD001700: N38-H138	BLAST_PRODOM
					IDURONATE 2SULFATASE PRECURSOR HYDROLASE SIGNAL GLYCOPROTEIN LYSOSOME MUCOPOLYSACCHARIDOSIS DISEASE MUTATION PD151796: M1-L37	BLAST_PRODOM
					SULFATASES DM01026 |P51689|38-562: N38-K135 |P51690|35-559: N38-L141 |P34059|28-486: N38-K135 |P08842|24-548: N38-K135	BLAST_DOMO

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					Sulfatases signature 1: A82-G94	MOTIFS
					Sulfatases signature 2: G127-H138	MOTIFS
16	7510298CD1	227	S71 T140 Y158	N184	Signal Peptide: M1-T30	HMMER
					Acyltransferase: I54-G227	HMMER_PFAM
					Cytosolic domain: T36-G227 Transmembrane domain: L15-W35 Non-cytosolic domain: M1-P14	TMHMMER
					Tafazzin signature PR00979: L63-L82, R94-T105, L108-R123, N170-N184	BLIMPS_PRINTS
					ZK809.2 PROTEIN PD107313: N40-P120	BLAST_PRODOM
					CHROMOSOME XVI COSMID TAFAZZIN ALTERNATIVE SPLICING TRANSMEMBRANE ZK809.2 PROTEIN PD137665: G26-Q129	BLAST_PRODOM
17	7510299CD1	388	S57 S76 S105 S150 S171 S231 S351 S377 T279	N5 N87 N103	signal_cleavage: M1-S23	SPSCAN
					Signal Peptide: M1-C17, M1-S21, M1-S23, M1-Y25, M1-A27, M1-G30	HMMER
					Cytosolic domain: M1-G6 Transmembrane domain: T7-S29 Non-cytosolic domain: G30-W388	TMHMMER
					UDPGLCNAC: A1,3-D-MANNOSIDE B-1,4-N-ACETYLGLUCOSAMINYLTRANSFERASE IV EC 2.4.1.145 ALPHA-1,3-MANNOSYLGLYCOPROTEIN BETA-1,4-N-ACETYLGLUCOSAMINYLTRANSFERASE N-GLYCOSYL OLIGOSACCHARIDE GLYCOPROTEIN PD185013:M1-A347	BLAST_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
18	7510368CD1	81	S36 S45		signal_cleavage: M1-A17	SPSCAN
					Signal Peptide: M1-A24	HMMER
					PROTOPORPHYRINOGEN; OXIDASE; DM03709|P50336|3-340: R3-L78	BLAST_DOMO
19	7510369CD1	412	S36 S45 S179 S228 S271 S275 S337 S398 T193 T245 T366		Signal Peptide: M1-A24	HMMER
					Flavin-containing amine oxidoreductase: I12-G412	HMMER_PFAM
					proto_IX_ox: protoporphyrinogen oxid: M1-T410	HMMER_TIGRFAM
					OXIDASE PROTOPORPHYRINOGEN OXIDOREDUCTASE PPO PORPHYRIN BIOSYNTHESIS FLAVOPROTEIN FAD HEME MITOCHONDRION PD004706: I53-R207 W227-V367	BLAST_PRODOM
					PROTOPORPHYRINOGEN; OXIDASE; DM03709 |P50336|3-340: R3-D340 |Q10062|1-353: T4-S337 |P32397|6-344: V5-D340 |P40012|12-390: V5-S196 E222-S338	BLAST_DOMO
					Cell attachment sequence: R88-D90	MOTIFS
20	7510377CD1	324	S31 S105 S160 S184 S238 T5 T122 T126 T219	N208 N214 N296	signal_cleavage: M1-G25	SPSCAN
					Signal Peptide: M1-E21, M1-A22, M1-G24, M1-S28, M1-G25	HMMER

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
21	7510026CD1	1041	S22 S40 S188 S297 S326 S368 S381 S398 S441 S561 S647 S690 S717 S736 S750 S788 S820 S879 S943 S1034 T25 T60 T67 T180 T388 T540 T567 T580 T692 T716 T727 T806 T845 T962 T1011 Y83	N438 N443 N515 N958 N1032	Signal Peptide: M1-A21, M1-G23	HMMER
					Protein kinase C conserved region 2 (CalB): T549-L648	HMMER_SMRT
					Phospholipase C, catalytic domain (part);: Q160-K310	HMMER_SMRT
					Phospholipase C, catalytic domain (part);: L412-R528	HMMER_SMRT
					C2 domain: C550-I633	HMMER_PFAM
					Phosphatidylinositol-specific phospholipase: E161-R311	HMMER_PFAM
					Phosphatidylinositol-specific phospholipase: L412-R528	HMMER_PFAM
					Phosphatidylinositol-specific phospholipase X-box domain proteins prof BL50007: L166-G211, C227-Q264, L295-R311, Y461-D502, A620-L656	BLIMPS_BLOCKS
					C2 domain signature PR00360: S561-M573, N593-F606, L617-N625	BLIMPS_PRINTS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
21, cont.					Phospholipase C signature PR00390: P165-Q183, E191-G211, A294-R311, M466-W487, W487-M505, L634-R644	BLIMPS_PRINTS
					PHOSPHOLIPASE C PHOSPHODIESTERASE HYDROLASE 1-PHOSPHATIDYLINOSITOL-4,5-BISPHOSPHATE LIPID DEGRADATION TRANSDUCER PHOSPHOINOSITIDE-SPECIFIC PD001214: E161-R311	BLAST_PRODOM
					BETA PHOSPHOLIPASE 1-PHOSPHATIDYLINOSITOL-4,5-BISPHOSPHATE PHOSPHODIESTERASE PLC154 HYDROLASE LIPID DEGRADATION TRANSDUCER PD023749: E315-Y411	BLAST_PRODOM
					PHOSPHOLIPASE 1-PHOSPHATIDYLINOSITOL-4,5-BISPHOSPHATE PHOSPHODIESTERASE BETA HYDROLASE LIPID DEGRADATION TRANSDUCER PLC154 PD011437: L900-K996	BLAST_PRODOM
					PHOSPHOLIPASE PHOSPHODIESTERASE C HYDROLASE 1-PHOSPHATIDYLINOSITOL-4,5-BISPHOSPHATE LIPID DEGRADATION TRANSDUCER PHOSPHOINOSITIDE-SPECIFIC PD001202: L412-R528	BLAST_PRODOM
					1-PHOSPHATIDYLINOSITOL-4,5-BISPHOSPHATE PHOSPHODIESTERASE D DM00855|A48047|58-521: S15-S368	BLAST_DOMO

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					1-PHOSPHATIDYLINOSITOL-4,5-BISPHOSPHATE PHOSPHODIESTERASE D DM00712 |A48047|523-820: A369-L667 |A53766|83-369: Y401-L667 |P13217|514-804: Q393-L667	BLAST_DOMO
					ATP/GTP-binding site motif A (P-loop): G181-S188	MOTIFS
22	7509168CD1	32			Glutathione S-transferase, N-terminal domain: K4-E29	HMMER_PFAM
23	7500607CD1	128	S38		I CARNITINE TRANSFERASE OPALMITOYLTRANSFERASE MITOCHONDRIAL ISOFORM CPT ACYLTRANSFERASE MITOCHONDRION OUTER PD014480: M1-V84	BLAST_PRODOM
24	7506079CD1	362	S20 S94 S111 S127 S335 T37 T187 T340		Conserved carboxylase domain: V45-N253	HMMER_PFAM
					Biotin-requiring enzyme: G294-I361	HMMER_PFAM
					Biotin-requiring enzymes attachment site: V304-G354	PROFILESCAN
					PYRUVATE CARBOXYLASE BIOTIN LIGASE PYRUVIC MULTIFUNCTIONAL ENZYME GLUCONEOGENESIS PCB ATP-BINDING PD002904: L32-K245	BLAST_PRODOM
					PYRUVATE CARBOXYLASE PRECURSOR PYRUVIC PCB LIGASE MULTIFUNCTIONAL ENZYME BIOTIN MANGANESE PD023149: M1-P36	BLAST_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					BIOTIN-REQUIRING ENZYMES ATTACHMENT SITE DM00738 |P11498|1030-1177: L214-E362 |P11154|1020-1168: L214-E362	BLAST_DOMO
					PYRUVATE CARBOXYLASE DM08298|P11498|915-1028: G99-P213	BLAST_DOMO
					PYRUVATE CARBOXYLASE DM03808|P11498|559-913: T37-V98	BLAST_DOMO
					Biotin-requiring enzymes attachment site: G318-S335	MOTIFS
25	7509259CD1	185	S81	N139	signal_cleavage: M1-G26	SPSCAN
					Signal Peptide: M1-G26	HMMER
					Iodothyronine deiodinase: L8-H184	HMMER_PFAM
					Cytosolic domain: D35-S185 Transmembrane domain: V15-P34 Non-cytosolic domain: M1-W14	TMHMMER
					Iodothyronine deiodinases proteins BL01205: I41-K68, D70-W99, P128-E181	BLIMPS_BLOCKS
					Uteroglobin family signatures: K27-N56	PROFILESCAN
					TYPE DEIODINASE 5'-DEIODINASE DI-OXIDOREDUCTASE HYDROLASE SELENIUM SELENOCYSTEINE IODOTHYRONINE TRANSMEMBRANE PD005013: L8-H184	BLAST_PRODOM
					DM02779 |P24389|4-256: L8-L183 |P49894|4-243: P4-K47, G49-S185, Q11-V133 |P55073|26-277: I31-K182 |P49898|25-268: G49-K182	BLAST_DOMO
26	7509263CD1	97	S29 S43 S56 T37		Eukaryotic-type carbonic anhydrase: W5-A77	HMMER_PFAM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					Eukaryotic-type carbonic anhydrases proteins BL00162: W16-P46, Y51-S73	BLIMPS_BLOCKS
					CARBONIC ANHYDRASE DEHYDRATASE LYASE CARBONATE ZINC PRECURSOR SIGNAL PROTEIN GLYCOPROTEIN PD000865: H4-G81	BLAST_PRODOM
					CARBONIC ANHYDRASE DM00356 |P00918|23-258: K24-G81 |P00919|23-258: G25-E84 |P00920|23-258: G25-G81 |P07630|23-258: G25-C94	BLAST_DOMO
27	7509360CD1	217	S41 S44 S53 T71	N193	Cytochrome P450: I51-V209	HMMER_PFAM
					Mitochondrial P450 signature PR00408: W126-L141, K142-V152	BLIMPS_PRINTS
					CYTOCHROME P450 CHOLESTEROL SIDECHAIN CLEAVAGE 11A1 MITOCHONDRIAL PRECURSOR CYPXIA1 P450SCC PD150160: M1-S53	BLAST_PRODOM
					CYTOCHROME P450 DM00022 |P05108|94-506: G94-K208 |P14137|92-503: G94-K208 |P00189|94-505: G94-K208 |Q07217|91-502: L93-K208	BLAST_DOMO
28	7509394CD1	189	S7 S27 S66 S71 S127 T14 T79 T93 T142 T162 Y108		Flavodoxin: N82-G175	HMMER_PFAM
					Cytosolic domain: R48-V189 Transmembrane domain: L25-F47 Non-cytosolic domain: M1-S24	TMHMMER

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
	--				Flavodoxin signature: N82-S126	PROFILESCAN
					Flavodoxin signature PR00369: I84-F97, V137-P148	BLIMPS_PRINTS
					NADPH CYTOCHROME P450 OXIDOREDUCTASE NADP REDUCTASE CPR FLAVOPROTEIN FMN FAD ENDOPLASMIC PD010852: E17-N82	BLAST_PRODOM
					FMN FLAVOPROTEIN OXIDOREDUCTASE REDUCTASE SYNTHASE NADP ELECTRON TRANSPORT P450 FAD PD000452: I83-W158	BLAST_PRODOM
					FLAVODOXIN DM00193 |P37039|77-188: R81-G175 |Q07994|74-186: G80-G175 |S38427|63-184: G80-V164 |P50126|58-168: R81-V184	BLAST_DOMO
29	7581076CD1	189	S71 S78 S133 S149 S168 S183 T113		lactate/malate dehydrogenase, NAD binding: L73-I127, L11-E37	HMMER_PFAM
					L-lactate dehydrogenase proteins BL00064: K48-Q95, S103-Q147, T148-T177	BLIMPS_BLOCKS
					L-lactate dehydrogenase active site: L136-Q181	PROFILESCAN
					L-lactate dehydrogenase signature PR00086: L100-P120, V124-N142, W154-W167	BLIMPS_PRINTS
					DEHYDROGENASE OXIDOREDUCTASE NAD MALATE L-LACTATE GLYCOLYSIS ACID TRICARBOXYLIC CYCLE MULTIGENE PD000350: L12-S168	BLAST_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					L-LACTATE DEHYDROGENASE DM00253 |I62761|19-331: L11-G169 |P04642|18-330: L11-G169 |P07864|17-329: L11-G169 |P33571|20-332: L11-G169	BLAST_DOMO
30	7504551CD1	73			Signal_cleavage: M1-R19	SPSCAN
					Short-chain dehydrogenases/reductases family proteins BL00061: G12-G21	BLIMPS_BLOCKS
					NADH-UBIQUINONE OXIDOREDUCTASE 49 KD SUBUNIT PRECURSOR EC 1.6.5.3 1.6.99.3 COMPLEX I49KD CI49KD NAD UBIQUINONE MITOCHONDRION IRON SULFUR 4FE4S TRANSIT PEPTIDE PD165511: M1-R32	BLAST_PRODOM
31	7500652CD1	115	S6 S7 S109 T12		H19N07.4 PROTEIN ACAT RELATED GENE PRODUCT PD155879: A48-G101	BLAST_PRODOM
					ACAT RELATED GENE PRODUCT 1 PD180603: P16-P47	BLAST_PRODOM
32	7500900CD1	403	S14 S293 T80 T90 T143 T230 T264 T319 T384 Y183		Aldehyde dehydrogenase family: M1-G403	HMMER_PFAM
					Aldehyde dehydrogenases glutamic acid proteins BL00687: G15-R32, Q100-E141, Y154-G190, G209-N255, T310-L359	BLIMPS_BLOCKS
					Aldehyde dehydrogenases active sites: T187-K233	PROFILESCAN
					Aldehyde dehydrogenases active sites: C214-V265	PROFILESCAN

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					DEHYDROGENASE OXIDOREDUCTASE ALDEHYDE NAD PROTEIN CLASS SEMIALDEHYDE PRECURSOR TRANSIT PEPTIDE PD000218: V5-K299, G165-G401	BLAST_PRODOM
					ALDEHYDE DEHYDROGENASE LIKE PROTEIN PD065537: Y246-G403	BLAST_PRODOM
					ALDEHYDE DEHYDROGENASES GLUTAMIC ACID DM00100|P51648|1-405: M1-G403	BLAST_DOMO
					ALDEHYDE DEHYDROGENASES GLUTAMIC ACID DM00100|P30838|1-408: V5-G403	BLAST_DOMO
					ALDEHYDE DEHYDROGENASES GLUTAMIC ACID DM00100|P43353|1-408: V5-G403	BLAST_DOMO
					ALDEHYDE DEHYDROGENASES GLUTAMIC ACID DM00100|P39616|1-409: F12-G403	BLAST_DOMO
					Aldehyde dehydrogenases cysteine active site: Y234-D245	MOTIFS
					Aldehyde dehydrogenases glutamic acid active site: L206-P213	MOTIFS
33	7501398CD1	165	S113 T123 T137		N-ACETYL-GLUCOSAMINE 6-PHOSPHATE DEACETYLASE GLCNAC 6P HYDROLASE CARBOHYDRATE METABOLISM NAGA PUTATIVE PROTEIN PD008493: F15-R136	BLAST_PRODOM
					DEACETYLASE; ACETYLGLUCOSAMINE; ACETYLGALACTOSAMINE; DM03012|P34480|64-417: D57-A135	BLAST_DOMO
34	7501417CD1	236	S45 S65 S139 S224 T21 T60 T144			

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
35	7501472CD1	454	S2 S139 S180 S226 S242 S337 S435 S440 T165 T189 T400 T445 Y204 Y317 T435	N8 N31 N32	PROLINE DEHYDROGENASE OXIDOREDUCTASE OXIDASE PROTEIN PRECURSOR DELTA1 PYRROLINE-5-CARBOXYLATE PUTA MITOCHONDRION TRANSIT PD003368: L135-L249, D272-L451	BLAST_PRODOM
					Prenyl group binding site (CAAX box): C452-V454	MOTIFS
36	7501489CD1	77			Signal_cleavage: M1-C17	SPSCAN
					Signal Peptide: M1-Q19	HMMER
					Signal Peptide: M1-A23	HMMER
					Signal Peptide: M1-F22	HMMER
					FLAP/GST2/LTC4S family proteins BL01297: G27-V40	BLIMPS_BLOCKS
					Nerve growth factor family signature: L7-R75	PROFILESCAN
					TRANSMEMBRANE LEUKOTRIENE BIOSYNTHESIS PROTEIN TRANSFERASE 5-LIPOXYGENASE ACTIVATING FLAP MK886-BINDING MICROSOMAL PD009387: M1-A52	BLAST_PRODOM
37	7501555CD1	248	S24 S174 S241 T100 T201		Zinc-binding dehydrogenase: P9-P248	HMMER_PFAM
					Zinc-containing alcohol dehydrogenases proteins BL00059: E29-M45, S58-T85, A88-F129, D142-A188	BLIMPS_BLOCKS
					Quinone oxidoreductase / zeta-crystallin proteins BL01162: E29-Y51, A119-R162	BLIMPS_BLOCKS
					OXIDOREDUCTASE ZINC DEHYDROGENASE ALCOHOL NAD PROTEIN FAMILY MULTIGENE NADP FORMALDEHYDE PD000104: G10-N191	BLAST_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					ZINC-CONTAINING ALCOHOL DEHYDROGENASES DM00064|P28304|3-297: V4-T205	BLAST_DOMO
					ZINC-CONTAINING ALCOHOL DEHYDROGENASES DM00064|Q02251|1415-1736: L2-E199	BLAST_DOMO
					ZINC-CONTAINING ALCOHOL DEHYDROGENASES DM00064|P43903|13-297: P12-T205	BLAST_DOMO
					ZINC-CONTAINING ALCOHOL DEHYDROGENASES DM00064|P11415|9-326: G10-K203	BLAST_DOMO
38	7501561CD1	76	S33			
39	7506108CD1	168	S74 T26 T91		CDP-alcohol phosphatidyltransferase: I6-M156	HMMER_PFAM
					Cytosolic domains: M53-A95, A153-K168; Transmembrane domains: S30-S52, L96-F115, G130-A152; Non-cytosolic domains: M1-C29, S116-M129	TMHMMER
					CDP-ALCOHOL PHOSPHATIDYLTRANSFERASES SIMILAR DM06665|Q10153|1-250: T16-K168	BLAST_DOMO
					CDP-ALCOHOL PHOSPHATIDYLTRANSFERASES SIMILAR DM06665|P06197|1-219: E4-G15, G19-K167	BLAST_DOMO
40	7506123CD1	187	S38 S43 S89 S148 T94 T167	N54 N114	Aminotransferases class-V pyridoxal-phosphate attachment site: G129-K179	PROFILESCAN
					AMINOTRANSFERASE BIOSYNT PD02379: M42-G80, V110-I124, L149-F181	BLIMPS_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					PHOSPHOSERINE AMINOTRANSFERASE DM02556|P10658|1-369: I65-L187, M1-L64, A141-K173	BLAST_DOMO
					PHOSPHOSERINE AMINOTRANSFERASE DM02556|P33330|1-394: N9-L64, D48-E183	BLAST_DOMO
					PHOSPHOSERINE AMINOTRANSFERASE DM02556|P44336|1-361: Q6-L63, Y66-F181	BLAST_DOMO
					PHOSPHOSERINE AMINOTRANSFERASE DM02556|P23721|1-360: Q6-D100, G13-H185	BLAST_DOMO
41	7506248CD1	1091	S2 S114 S263 S455 S459 S510 S564 S567 S571 S779 S789 S833 S1006 T10 T198 T291 T353 T596 T686 T920 T1005 T1034 Y672 Y844 Y1083	N75 N346 N440	CoA binding domain: L482-K606	HMMER_PFAM
					CoA-ligase: M632-L783	HMMER_PFAM
					sucCoAalpha: succinyl-CoA synthetase: T481-Y791	HMMER_TIGRFAM
					LYASE PROTEIN ATPCITRATE PROSLYASE CITRATE CLEAVAGE ENZYME LIPID SYNTHESIS ATPBINDING PD010086: M1-D257, I236-L419	BLAST_PRODOM
					LYASE ATP-CITRATE PROSLYASE CITRATE CLEAVAGE ENZYME LIPID SYNTHESIS PHOSPHORYLATION ATP-BINDING PD014373: G420-D634	BLAST_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					LYASE ATP-CITRATE PROSLYASE CITRATE CLEAVAGE ENZYME LIPID SYNTHESIS PHOSPHORYLATION ATP-BINDING PD013870: Y791-F879	BLAST_PRODOM
					CITRATE LYASE SYNTHASE TRICARBOXYLIC ACID CYCLE ENZYME ALLOSTERIC PRECURSOR TRANSIT PD000990: I880-K1067	BLAST_PRODOM
					CITRATE; ATP; DM07845|P53396|1-517: M1-D504	BLAST_DOMO
					CITRATE; ATP; DM07844|P53396|807-1104: D793-M1091	BLAST_DOMO
					SUCCINATE--COA LIGASE (ADP-FORMING) ALPHA CHAIN DM01167|P53396|519-805: F505-E792	BLAST_DOMO
					CITRATE; ATP; DM07845|P53585|1-513: M1-D504	BLAST_DOMO
					Leucine zipper pattern: L55-L76	MOTIFS
					ATP-citrate lyase / succinyl-CoA ligases family signature 1: S651-D680	MOTIFS
					ATP-citrate lyase / succinyl-CoA ligases family active site: G736-G752	MOTIFS
					ATP-citrate lyase / succinyl-CoA ligases family signature 3: G273-G297	MOTIFS
42	7506347CD1	473	S21 S136 S243 S258 S450 T33 T48 T70 T235 T257 T375 T431 T436 Y170	N255 N256	Flavin-binding monooxygenase-like: A6-L473	HMMER_PFAM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					Pyridine nucleotide-disulphide oxidoreductases class-II active site: S100-L153	PROFILESCAN
					Flavin-containing monooxygenase (FMO) signature PR00370: R8-K23, E31-T49, F81-P97, K126-A144, P262-D280, G317-G336, F414-R435	BLIMPS_PRINTS
					Flavin-containing amine oxidase signature PR00757: R8-E27	BLIMPS_PRINTS
					DIMETHYLANILINE MONOOXYGENASE (N-OXIDE-FORMING) DM02407|Q01740|1-334: A6-V51, R46-P277, V286-I297	BLAST_DOMO
					DIMETHYLANILINE MONOOXYGENASE (N-OXIDE-FORMING) DM02585|Q01740|336-530: F278-L473	BLAST_DOMO
					DIMETHYLANILINE MONOOXYGENASE (N-OXIDE-FORMING) DM02407|P36366|1-334: A6-V51, F35-P277, V286-I297	BLAST_DOMO
					DIMETHYLANILINE MONOOXYGENASE (N-OXIDE-FORMING) DM02407|P31513|1-334: K7-T48, F35-P277, E287-F298	BLAST_DOMO
					Lipocalin signature: E36-T48	MOTIFS
43	7509172CD1	139	S23 S95 S103 S131	N56	Signal_cleavage: M1-R21	SPSCAN
					Signal Peptide: M1-S18, M1-R21, M1-Q22, M1-R25	HMMER
					Cytosolic domain: Q22-A139; Transmembrane domain: F4-R21; Non-cytosolic domain: M1-P3	TMHMMER

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					CYTOCHROME P450 MONOOXYGENASE OXIDOREDUCTASE HEME ELECTRON TRANSPORT MEMBRANE MICROSOME ENDOPLASMIC PD000021: P30-P101	BLAST_PRODOM
					CYTOCHROME P450 DM00022|I48164|11-474: L11-G111	BLAST_DOMO
					CYTOCHROME P450 DM00022|P10632|48-474: K48-G111	BLAST_DOMO
					CYTOCHROME P450 DM00022|A46588|15-478: V8-G111	BLAST_DOMO
					CYTOCHROME P450 DM00022|P33262|48-474: K48-G111	BLAST_DOMO
44	7510421CD1	66	S33 S53		Signal_cleavage: M1-A16	SPSCAN
					DELTA 1-PYRROLINE-5-CARBOXYLATE REDUCTASE DM00944|P32322|1-270: M1-L65	BLAST_DOMO
45	7504625CD1	140	T64 T76		Signal_cleavage: M1-A23	SPSCAN
					Short-chain dehydrogenases/reductases family proteins BL00061: G81-G91	BLIMPS_BLOCKS
					Alcohol dehydrogenase superfamily signature PR00080: G81-H92	BLIMPS_PRINTS
					Glucose/ribitol dehydrogenase family signature PR00081: V11-A28, G81-H92	BLIMPS_PRINTS
					SHORT-CHAIN ALCOHOL DEHYDROGENASE FAMILY DM00034|P50160|50-326: R6-G91	BLAST_DOMO
					SHORT-CHAIN ALCOHOL DEHYDROGENASE FAMILY DM00034|P50197|1-246: A8-E104	BLAST_DOMO

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					SHORT-CHAIN ALCOHOL DEHYDROGENASE FAMILY DM00034|Q11020|2-244: K10-R94	BLAST_DOMO
					SHORT-CHAIN ALCOHOL DEHYDROGENASE FAMILY DM00034|P08088|1-255: R6-G91	BLAST_DOMO
46	7504776CD1	356	S65 S94 S193 S228 S336 T143 T191 T204		Signal_cleavage: M1-P18	SPSCAN
					Signal Peptide: M1-P18	HMMER
					Signal Peptide: M1-C22	HMMER
					Signal Peptide: M1-G24	HMMER
					Cytochrome P450: P28-I341	HMMER_PFAM
					Cytochrome P450 cysteine heme-iron ligand proteins BL00086: Y280-F311	BLIMPS_BLOCKS
					Cytochrome P450 cysteine heme-iron ligand signature: F260-D312	PROFILESCAN
					Mitochondrial P450 signature PR00408: A150-L167, H168-D181, N196-V214, F260-P268, L281-C290, C290-F301	BLIMPS_PRINTS
					E-class P450 group I signature PR00463: Q57-V76, L81-F102, H139-T156, V159-G185, E202-P220, N243-N267, Y280-C290, C290-L313	BLIMPS_PRINTS
					E-class P450 group II signature PR00464: H139-L167, H168-G185, R197-M217, G237-K252, E253-P268, S277-C290, C290-L313	BLIMPS_PRINTS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					CYTOCHROME P450 MONOOXYGENASE OXIDOREDUCTASE HEME ELECTRON TRANSPORT MEMBRANE MICROSOME ENDOPLASMIC PD000021: Q80-D226, H248-P322, V184-L266, L34-R96	BLAST_PRODOM
					CYTOCHROME P450 DM00022|P05093|69-479: G69-M99, E100-P328	BLAST_DOMO
					CYTOCHROME P450 DM00022|S52756|69-479: G69-P97, E100-P328	BLAST_DOMO
					CYTOCHROME P450 DM00022|P05185|69-479: G69-M99, E100-G326	BLAST_DOMO
					CYTOCHROME P450 DM00022|P11715|69-478: G69-M99, E100-P328	BLAST_DOMO
					Cytochrome P450 cysteine heme-iron ligand signature: F283-G292	MOTIFS
47	7504927CD1	141	T89 Y73		Signal_cleavage: M1-Q13	SPSCAN
					NADH-UBIQUINONE OXIDOREDUCTASE B22 UBIQUINONE SUBUNIT COMPLEX IB22 CIB22 NAD MITOCHONDRION PD034663: L10-Y97, E92-P111	BLAST_PRODOM
					NADH-UBIQUINONE OXIDOREDUCTASE B22 SUBUNIT EC 1.6.5.3 1.6.99.3 COMPLEX IB22 CIB22 NAD UBIQUINONE MITOCHONDRION ACETYLATION PD095720: T113-M141	BLAST_PRODOM
48	7505010CD1	398	S83 S239 S256 S300 S385 T104 T186 T212 T225 T335	N210 N237	Signal Peptide: M1-A30	HMMER

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
48, cont.					AIR carboxylase: C292-I393	HMMER_PFAM
					SAICAR synthetase: I35-A279	HMMER_PFAM
					purC: phosphoribosylaminoimidazole-succinoc: K37-E262	HMMER_TIGRFAM
					purE: phosphoribosylaminoimidazole carboxyl: V294-R394	HMMER_TIGRFAM
					Glycosyl hydrolases family 10 proteins BL00591: Y148-A160, G180-I190, I229-W240	BLIMPS_BLOCKS
					SAICAR synthetase proteins BL01057: V57-F91, M119-S133, G140-D159, N210-D223, I229-W243, Q250-K261	BLIMPS_BLOCKS
					AIR carboxylase. PF00731: V294-E322, V325-A361	BLIMPS_PFAM
					SYNTHASE SAICAR SYNTHETASE PURINE BIOSYNTHESIS LIGASE PHOSPHORIBOSYLAMINOIMIDAZOLESUCCINOCARBOXAMIDE PROTEIN PHOSPHORIBOSYLAMIDOIMIDAZOLESUCCINOCARBOXAMIDE MULTIFUNCTIONAL PD003043: K37-S133, K106-E276	BLAST_PRODOM
					SAICAR SYNTHETASE DM01073|P38024|8-236: L33-E262	BLAST_DOMO
					SAICAR SYNTHETASE DM01073|Q10457|11-238: L39-E262	BLAST_DOMO
					AIR CARBOXYLASE DOMAIN DM07131|P38024|238-425: V263-V350, G344-L398	BLAST_DOMO

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					SAICAR synthetase signature 2: L213-G221	MOTIFS
49	7505173CD1	115			Signal_cleavage: M1-C37	SPSCAN
					Signal Peptide: M1-A18	HMMER
					Phosphopantetheine attachment site: Q42-I107	HMMER_PFAM
					acyl_carrier: acyl carrier protein: T41-K111	HMMER_TIGRFAM
					Phosphopantetheine attachment site: C37-V104	PROFILESCAN
					Phosphopantetheine attachment PF00550:D66-M79	BLIMPS_PFAM
					ACYL CARRIER PROTEIN, MITOCHONDRIAL PRECURSOR ACP NADH-UBIQUINONE OXIDOREDUCTASE 9.6 KD SUBUNIT EC 1.6.5.3 1.6.99.3 CISDAP FATTY ACID BIOSYNTHESIS PHOSPHOPANTETHEINE MITOCHONDRION TRANSIT PD169104: M25-L57	BLAST_PRODOM
					PHOSPHOPANTETHEINE ATTACHMENT SITE DM00046|P32463|35-121: P50-A108	BLAST_DOMO
					PHOSPHOPANTETHEINE ATTACHMENT SITE DM00046|A26935|1-75: Q56-I107	BLAST_DOMO
					PHOSPHOPANTETHEINE ATTACHMENT SITE DM00046|P11943|47-130: L57-I107	BLAST_DOMO
					EF-hand calcium-binding domain: D91-I103	MOTIFS
					Phosphopantetheine attachment site: D66-M81	MOTIFS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
50	7510061CD1	744	S41 S108 S215 S235 S244 S358 S456 S507 S637 T80 T126 T301 T364 T366 T455 T638 T678 T721 T735 Y312	N233 N350 N400 N666	Signal_cleavage: M1-G32	SPSCAN
					Cytosolic domain: M1-R18; Transmembrane domain: L19-S41; Non-cytosolic domain: P42-V744	TMHMMER
					SULFATE TRANSFERASE N-DEACETYLASE/N-SULFOTRANSFERASE SULFOTRANSFERASE GLUCOSAMINYL HEPARAN N HSST DEACETYLASE/N-SULFOTRANSFERASE TRANSMEMBRANE PD011877: V87-R602	BLAST_PRODOM
					HEPARIN SULFATE N-DEACETYLASE/N-SULFOTRANSFERASE N HSST N-HEPARIN SULFOTRANSFERASE GLUCOSAMINYL DEACETYLASE/N-SULFOTRANSFERASE TRANSFERASE PD043426: M1-V86	BLAST_PRODOM
					SULFATE TRANSFERASE HEPARAN GLUCOSAMINYL N-DEACETYLASE/N-SULFOTRANSFERASE SULFOTRANSFERASE N HSST DEACETYLASE/N-SULFOTRANSFERASE TRANSMEMBRANE PD007660: L607-Q714	BLAST_PRODOM
					SIGNAL-ANCHOR TRANSMEMBRANE DM07899|P52850|1-882: M1-Q714	BLAST_DOMO

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					SIGNAL-ANCHOR TRANSMEMBRANE DM07899|P52848|1-881: R9-Q714	BLAST_DOMO
51	7510091CD1	138	S46		Signal_cleavage: M1-G44	SPSCAN
					GLYCOSYLTRANSFERASE TRANSFERASE ALG2 SIMILAR GLYCOPROTEIN TRANSMEMBRANE M MUSCULUS MER5 OTHER PD011566: V17-L122	BLAST_PRODOM
52	7510109CD1	930	S181, S353	N644	Acyl-CoA dehydrogenase, C-terminal domain: G793-L921	HMMER_PFAM
					Acyl-CoA dehydrogenase, middle domain: A682-V789	HMMER_PFAM
					haloacid dehalogenase-like hydrolase: Y42-P234	HMMER_PFAM
					Acyl-CoA dehydrogenases proteins BL00072: Y707-I719, K756-F796, H810-R860	BLIMPS_BLOCKS
					Acyl-CoA dehydrogenases signatures: E685-P738	PROFILESCAN
					Haloacid dehalogenase/epoxide hydrolase family signature PR00413: Y42-L53, R512-Y525, I145-N158, F175-P191, I193-D213, E220-D233	BLIMPS_PRINTS
					PROTEIN DEHYDROGENASE ACYLCOA OXIDOREDUCTASE FLAVOPROTEIN FAD OXIDASE FATTY ACID METABOLISM PD000396: R566-G771, P792-V903	BLAST_PRODOM
					SIMILAR TO ACYLCOA DEHYDROGENASES AND EPOXIDE HYDROLASES HYDROLASE PD120138: H392-L474	BLAST_PRODOM
					ACYL-COA DEHYDROGENASES DM00853|I49605|32-408: E591-V903	BLAST_DOMO

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					ACYL-COA DEHYDROGENASES DM00853|P16219|32-408: E591-V903	BLAST_DOMO
					ACYL-COA DEHYDROGENASES DM00853|Q06319|3-383: R658-V903	BLAST_DOMO
					ACYL-COA DEHYDROGENASES DM00853|P45954|55-430: E571-Q902	BLAST_DOMO
					Eukaryotic thiol (cysteine) proteases histidine active site: I808-S818	MOTIFS
53	7510121CD1	260	S106 S233 S240 S251 T57	N174	Signal_cleavage: M1-A20	SPSCAN
					Signal Peptide: M5-P22	HMMER
					Signal Peptide: M5-K26	HMMER
					Signal Peptide: M1-R25	HMMER
					Signal Peptide: M1-K26	HMMER
					short chain dehydrogenase: P40-L260	HMMER_PFAM
					Alcohol dehydrogenase superfamily signature PR00080: K119-V130, S170-L178, Y202-K221	BLIMPS_PRINTS
					Glucose/ribitol dehydrogenase family signature PR00081: V43-E60, K119-V130, L164-H180, Y202-K221	BLIMPS_PRINTS
					SHORT-CHAIN ALCOHOL DEHYDROGENASE FAMILY DM00034|S42651|28-318: V45-K221	BLAST_DOMO
					SHORT-CHAIN ALCOHOL DEHYDROGENASE FAMILY DM00034|S39394|69-356: V43-L178, L190-R219	BLAST_DOMO
					SHORT-CHAIN ALCOHOL DEHYDROGENASE FAMILY DM00034|P13653|69-356: V43-L178, L190-R219	BLAST_DOMO

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					SHORT-CHAIN ALCOHOL DEHYDROGENASE FAMILY DM00034|S30167|82-368: V44-L178, L190-R219	BLAST_DOMO
54	7510797CD1	714	S19 S43 S74 S180 S205 S230 S236 S246 S253 S346 S583 S664 S707 T149 T161 T299 T323 T368 T397 T425 T482 T525 T544 T636 T699 Y542	N117 N420 N445 N508	3'5'-cyclic nucleotide phosphodiesterase: Y443-Y694	HMMER_PFAM
					3'5'-cyclic nucleotide phosphodiesterases proteins BL00126: L402-H438, Y443-H454, L469-D509, T525-E563, D610-S664	BLIMPS_BLOCKS
					3'5'-cyclic nucleotide phosphodiesterases signature: V465-H516	PROFILESCAN
					3'5'-cyclic nucleotide phosphodiesterase signature PR00387: S439-V452, H484-A499, A511-K527, I606-D619, E623-E639	BLIMPS_PRINTS
					CAMP-SPECIFIC 3',5'CYCLIC PHOSPHODIESTERASE 8A EC 3.1.4.17 HYDROLASE CAMP PD185095: M1-Q231, V166-N441	BLAST_PRODOM
					PHOSPHODIESTERASE 3' HYDROLASE 5'CYCLIC CGMP CAMP ALTERNATIVE SPLICING CGMP-SPECIFIC MULTIGENE PD001130: N441-N693	BLAST_PRODOM

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
					3'5'-CYCLIC NUCLEOTIDE PHOSPHODIESTERASES DM00370|Q07343|316-709: I376-W697	BLAST_DOMO
					3'5'-CYCLIC NUCLEOTIDE PHOSPHODIESTERASES DM00370|P14645|95-473: D371-W697	BLAST_DOMO
					3'5'-CYCLIC NUCLEOTIDE PHOSPHODIESTERASES DM00370|P27815|343-722: D371-W697	BLAST_DOMO
					3'5'-CYCLIC NUCLEOTIDE PHOSPHODIESTERASES DM00370|I38416|167-546: I376-W697	BLAST_DOMO
					3'5'-cyclic nucleotide phosphodiesterases signature: H484-F495	MOTIFS
55	7504944CD1	600	S46 S82 S249 S403 S513 T192 T204 T226 T235 T472 Y64 Y584	N173 N272 N369 N580	Signal_cleavage: M1-G22	SPSCAN
					Signal Peptide: M1-G22	HMMER
					Signal Peptide: M1-Q24	HMMER
					Signal Peptide: M1-L28	HMMER
					Glycosyl hydrolases family 2, immunog: T226-D311	HMMER_PFAM
					Glycosyl hydrolases family 2, TIM bar: L292-N580	HMMER_PFAM
					Glycosyl hydrolases family 2, sugar b: G22-P224	HMMER_PFAM
					Cytosolic domain: M1-A6; Transmembrane domain: V7-Y29; Non-cytosolic domain: P30-T600	TMHMMER

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
55, cont.					Glycosyl hydrolases family 2 proteins BL00719: E32-W45, D207-I231, G273-K317, D318-E355, R385-E400, Q482-G497, N537-S566	BLIMPS_BLOCKS
					Glycosyl hydrolases family 2 signatures: N368-P424	PROFILESCAN
					Glycosyl hydrolase family 2 signature PR00132: A126-E141, D318-P336, D386-P401, K483-F498	BLIMPS_PRINTS
					Glycosyl hydrolase family 26 signature PR00739: G441-Y453	BLIMPS_PRINTS
					BETAGALACTOSIDASE HYDROLASE GLYCOSIDASE LACTASE BETAGLUCURONIDASE PRECURSOR LYSOSOME GLYCOPROTEIN SIGNAL LARGE PD002163: S35-D-311, K317-Y490, H375-P564	BLAST_PRODOM
					BETAGLUCURONIDASE PRECURSOR HYDROLASE GLYCOSIDASE LYSOSOME GLYCOPROTEIN SIGNAL BETAG1 MUCOPOLYSACCHARIDOSIS DISEASE PD009825: M1-P34	BLAST_PRODOM
					BETA-GLUCURONIDASE DM02380|P08236|77-650: M77-R306, Y291-T600	BLAST_DOMO
					BETA-GLUCURONIDASE DM02380|P06760|77-647: M77-L322, Y291-T600	BLAST_DOMO
					BETA-GLUCURONIDASE DM02380|P12265|77-647: M77-R306, Y291-T600	BLAST_DOMO
					BETA-GLUCURONIDASE DM02380|A32576|77-648: M77-R306, Y291-T600	BLAST_DOMO
					Glycosyl hydrolases family 2 signature 1: N328-I353	MOTIFS

Table 3

SEQ ID NO:	Incyte Polypeptide ID	Amino Acid Residues	Potential Phosphorylation Sites	Potential Glycosylation Sites	Signature Sequences, Domains and Motifs	Analytical Methods and Databases
55, cont.					Glycosyl hydrolases family 2 acid/base catalyst: D386-E400	MOTIFS

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
56/7509350CB1 1671	1-197, 1-214, 1-219, 1-249, 1-250, 1-260, 1-268, 1-271, 1-282, 1-289, 1-396, 1-427, 1-431, 1-457, 1-469, 1-470, 1-490, 1-598, 1-619, 1-634, 1-637, 1-638, 1-663, 1-801, 1-1661, 3-239, 3-634, 8-468, 8-508, 8-618, 8-713, 8-738, 8-739, 8-744, 8-745, 8-766, 8-828, 8-837, 8-952, 15-272, 16-682, 23-281, 23-722, 29-427, 95-604, 96-582, 132-259, 171-651, 235-495, 264-788, 277-427, 300-528, 300-916, 325-592, 326-427, 364-603, 418-637, 494-741, 497-993, 525-1373, 531-883, 533-701, 580-998, 588-698, 613-1373, 663-1373, 669-1373, 712-1373, 792-979, 802-1006, 802-1220, 802-1333, 812-1005, 951-1208, 951-1320, 1008-1231, 1008-1241, 1017-1501, 1017-1598, 1017-1616, 1017-1648, 1018-1365, 1018-1599, 1019-1372, 1019-1546, 1019-1648, 1019-1658, 1020-1632, 1020-1649, 1021-1622, 1025-1488, 1026-1615, 1036-1624, 1042-1598, 1046-1323, 1051-1616, 1056-1662, 1075-1662, 1087-1419, 1103-1598, 1105-1612, 1120-1616, 1150-1654, 1152-1657, 1156-1306, 1156-1662, 1182-1621, 1206-1613, 1217-1422, 1222-1662, 1237-1453, 1243-1662, 1246-1662, 1247-1662, 1250-1619, 1261-1662, 1262-1622, 1285-1518, 1319-1635, 1372-1636, 1382-1662, 1388-1663, 1388-1671, 1458-1662, 1477-1662, 1498-1662, 1500-1662, 1550-1635, 1550-1662
57/7509325CB1 2201	1-252, 1-790, 1-2092, 3-351, 4-684, 6-334, 6-339, 6-354, 13-199, 13-226, 14-236, 509-1087, 602-817, 625-1266, 689-1208, 721-1269, 723-945, 723-969, 723-1035, 723-1147, 765-1450, 770-1270, 770-1281, 806-1309, 807-1309, 841-1039, 893-1743, 900-1427, 900-1743, 909-1743, 912-1607, 912-1742, 921-1169, 921-1175, 927-1743, 928-1743, 970-1741, 982-1450, 983-1469, 984-1253, 988-1518, 989-1743, 993-1705, 998-1529, 1005-1652, 1018-1626, 1042-1689, 1066-1742, 1088-1345, 1091-1743, 1100-1441, 1100-1602, 1106-1371, 1106-1421, 1106-1457, 1106-1459, 1106-1560, 1106-1574, 1106-1584, 1106-1632, 1106-1672, 1106-1764, 1109-1624, 1119-1671, 1144-1652, 1152-1574, 1158-1271, 1158-1570, 1177-1430, 1178-1741, 1215-1711, 1280-1664, 1288-1518, 1332-2033, 1360-1970, 1361-1915, 1373-1638, 1402-1743, 1407-1865, 1408-1717, 1429-1869, 1430-1979, 1435-2108, 1445-1981, 1445-2003, 1456-2108, 1464-1992, 1478-1852, 1486-2034, 1503-2044, 1507-2034, 1528-2031, 1531-1766, 1531-2066, 1531-2070, 1533-2094, 1534-2038, 1535-2143, 1553-2058, 1553-2134, 1568-1827, 1568-2094, 1571-1790, 1591-2020, 1591-2096, 1596-2156, 1614-2071, 1622-2201, 1640-2198, 1643-2087, 1646-1849, 1647-1849, 1652-2094, 1659-2120, 1708-2094, 1729-2094, 1775-1926, 1786-2074, 1842-2096, 1862-2080

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
58/7509337CB1 1231	1-207, 1-280, 1-1203, 1-1220, 37-448, 49-271, 49-433, 62-327, 186-364, 186-434, 186-454, 186-977, 186-1024, 186-1036, 186-1046, 196-1035, 328-1047, 454-1043, 454-1044, 458-885, 460-960, 462-669, 474-713, 474-1175, 488-1064, 511-774, 514-748, 515-774, 530-1049, 543-1016, 547-839, 568-1038, 578-806, 581-864, 582-992, 663-1206, 667-910, 668-951, 711-1039, 718-1205, 725-1124, 728-1223, 729-1231, 739-1167, 752-1037, 754-1205, 754-1210, 757-1212, 759-1204, 765-1208, 766-1231, 795-1206, 796-1167, 796-1205, 796-1220, 804-1231, 808-1205, 813-1206, 827-1210, 829-1212, 835-1213, 847-1205, 852-1198, 860-1208, 869-1205, 901-992, 921-1188, 922-1212, 971-1205, 1049-1205, 1112-1211
59/7509353CB1 2002	1-257, 1-323, 1-352, 1-354, 1-421, 1-459, 1-461, 1-511, 1-528, 1-556, 1-557, 1-575, 1-618, 1-630, 1-688, 1-691, 1-1991, 11-199, 11-273, 38-266, 42-158, 55-342, 65-574, 86-644, 109-689, 137-536, 219-513, 219-690, 229-899, 240-465, 696-1348, 696-1352, 714-1552, 773-1068, 834-1667, 837-1663, 876-1142, 876-1355, 936-1228, 973-1703, 996-1845, 1003-1809, 1011-1488, 1037-1826, 1058-1976, 1067-1972, 1076-1826, 1077-1972, 1093-1971, 1094-1972, 1095-1972, 1099-1972, 1105-1362, 1116-1972, 1134-1972, 1140-1972, 1142-1972, 1144-1972, 1151-1793, 1152-1844, 1163-1995, 1174-1972, 1175-1972, 1195-1734, 1206-1972, 1210-1972, 1225-1971, 1227-1508, 1235-1972, 1249-1529, 1250-1690, 1250-1895, 1256-1972, 1267-1972, 1290-1975, 1329-1972, 1341-1695, 1359-1972, 1364-1677, 1366-1817, 1371-1972, 1388-1964, 1419-1591, 1419-1974, 1420-1976, 1426-1949, 1439-1961, 1441-1916, 1466-1953, 1481-1940, 1503-1932, 1514-1995, 1545-1995, 1552-1857, 1614-1884, 1620-1926, 1640-1906, 1698-1947, 1759-1996, 1759-2002, 1783-1942
60/7509354CB1 1657	1-174, 1-175, 1-1657, 2-153, 3-321, 10-377, 23-106, 23-154, 25-182, 35-183, 37-707, 39-331, 176-460, 186-472, 191-370, 194-463, 221-284, 221-497, 389-497, 481-1345, 498-1141, 498-1345, 500-1345, 505-1033, 509-1345, 511-1345, 512-1059, 513-1345, 517-1345, 522-1476, 537-1345, 546-1476, 555-1345, 556-1476, 558-800, 563-807, 571-1476, 573-732, 573-1476, 578-835, 578-1097, 592-1345, 593-1345, 601-1476, 615-1232, 626-1476, 627-1345, 636-930, 644-832, 647-1188, 649-1345, 650-1345, 653-1076, 664-1281, 666-1345, 670-1355, 681-1476, 682-1477, 684-1476, 684-1477, 687-1476, 689-1142, 693-1345, 694-1345, 695-1353, 707-1258, 713-1400, 715-1282, 720-1477, 721-1254, 721-1476, 726-1345, 729-1274, 737-1341, 742-1345, 765-1036, 769-1476, 790-891, 801-1345, 828-1345, 828-1482, 840-1504, 842-1345, 866-1142, 878-1157, 888-1488, 891-1457, 894-1179, 905-1481, 907-1206, 908-1175, 912-1452, 915-1216, 921-1082, 923-1345, 929-1116, 933-1179, 933-1555, 941-1314, 950-1170, 965-1193, 970-1197, 978-1455, 979-1280, 979-1281,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
60 - Cont'd	983-1224, 1002-1646, 1006-1478, 1010-1267, 1010-1597, 1036-1438, 1036-1532, 1036-1644, 1037-1644, 1043-1557, 1047-1259, 1050-1227, 1051-1307, 1063-1284, 1070-1621, 1071-1298, 1075-1200, 1075-1522, 1075-1575, 1076-1609, 1080-1540, 1098-1236, 1124-1639, 1140-1382, 1146-1657, 1148-1314, 1150-1430, 1184-1470, 1191-1386, 1194-1436, 1197-1646, 1213-1645, 1213-1646, 1223-1643, 1226-1642, 1230-1646, 1231-1643, 1233-1582, 1239-1642, 1244-1646, 1245-1643, 1249-1499, 1260-1653, 1264-1494, 1270-1570, 1275-1604, 1290-1495, 1297-1643, 1302-1547, 1307-1645, 1311-1646, 1343-1643, 1394-1657, 1404-1653, 1412-1657, 1419-1650, 1428-1656, 1456-1643, 1457-1657, 1467-1657, 1468-1657, 1519-1645
61/7509385CB1 1862	1-174, 1-175, 1-472, 1-704, 1-1827, 2-153, 23-106, 23-154, 25-182, 35-183, 37-851, 37-992, 39-425, 39-765, 39-769, 39-853, 39-855, 39-937, 267-499, 274-589, 275-588, 299-817, 320-903, 323-583, 324-603, 350-629, 377-659, 396-665, 396-697, 396-901, 396-905, 396-919, 396-982, 396-994, 397-642, 440-910, 446-654, 446-811, 458-701, 473-716, 476-988, 491-806, 500-1139, 540-727, 541-1220, 544-1510, 547-1510, 553-1510, 562-1041, 568-1510, 585-1510, 588-782, 603-1510, 618-1510, 624-1510, 631-1510, 640-1510, 650-1510, 653-1306, 657-1641, 668-1510, 670-1198, 676-1510, 677-1224, 678-1510, 682-1510, 702-1510, 711-1641, 720-1510, 721-1641, 723-965, 728-972, 736-1641, 738-897, 738-1641, 743-1000, 743-1262, 757-1510, 758-1510, 766-1641, 780-1397, 791-1641, 792-1510, 801-1095, 809-997, 812-1353, 814-1510, 815-1510, 818-1241, 829-1446, 831-1510, 835-1520, 846-1641, 847-1642, 849-1641, 849-1642, 852-1641, 854-1307, 858-1510, 859-1510, 860-1518, 872-1423, 878-1565, 880-1447, 885-1642, 886-1419, 886-1641, 891-1510, 894-1439, 902-1506, 907-1510, 930-1201, 934-1641, 955-1056, 966-1510, 993-1510, 993-1647, 1005-1669, 1007-1510, 1031-1307, 1043-1322, 1053-1653, 1056-1622, 1059-1344, 1070-1646, 1072-1371, 1073-1340, 1077-1617, 1080-1381, 1086-1247, 1088-1510, 1094-1281, 1098-1344, 1098-1720, 1106-1479, 1115-1335, 1130-1358, 1135-1362, 1143-1620, 1144-1445, 1144-1446, 1148-1389, 1167-1811, 1171-1643, 1175-1432, 1175-1762, 1201-1603, 1201-1697, 1201-1809, 1202-1809, 1208-1722, 1212-1424, 1215-1392, 1216-1472, 1228-1449, 1235-1786, 1236-1463, 1240-1365, 1240-1687, 1240-1740, 1241-1774, 1245-1705, 1263-1401, 1289-1804, 1305-1547, 1311-1829, 1313-1479, 1315-1595, 1349-1635, 1356-1551, 1359-1601, 1362-1811, 1378-1810, 1378-1811, 1388-1808, 1391-1807, 1395-1811, 1396-1808, 1398-1747, 1404-1807, 1409-1811, 1410-1808, 1414-1664, 1425-1818, 1429-1659, 1435-1735, 1440-1769, 1455-1660, 1462-1808, 1467-1712, 1472-1810, 1476-1827, 1508-1808, 1559-1835, 1569-1818, 1577-1862, 1584-1815, 1593-1821, 1621-1808, 1622-1834, 1632-1835, 1633-1829, 1684-1810
62/7509216CB1 1471	1-1471, 495-999, 822-1221

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
63/7509376CB1 1509	1-231, 1-249, 1-1509, 97-673, 109-357, 110-495, 214-484, 411-569, 491-1037, 860-1259
64/7501927CB1 1778	1-247, 1-248, 1-253, 1-289, 1-315, 1-455, 1-491, 1-558, 1-560, 1-576, 1-621, 1-705, 1-733, 1-1778, 2-449, 3-281, 10-560, 268-512, 308-368, 378-504, 392-489, 397-670, 423-490, 542-639, 1095-1761, 1285-1777, 1433-1769
65/7503274CB1 1393	1-111, 1-180, 1-189, 1-218, 1-221, 1-235, 1-245, 1-247, 1-251, 1-317, 1-385, 1-387, 1-466, 1-504, 1-597, 1-603, 1-608, 1-639, 1-653, 1-657, 1-850, 1-881, 1-905, 2-327, 3-231, 3-329, 4-308, 4-678, 5-205, 5-214, 5-562, 7-338, 7-1393, 37-355, 49-624, 50-308, 52-267, 104-369, 123-805, 125-331, 126-386, 144-894, 161-419, 182-906, 188-382, 192-459, 192-813, 245-487, 245-493, 245-503, 267-721, 286-868, 291-539, 291-568, 291-575, 298-904, 302-564, 302-579, 305-566, 323-986, 333-652, 334-563, 334-575, 336-578, 342-846, 360-611, 372-688, 376-617, 387-640, 390-604, 390-821, 464-726, 464-752, 474-1007, 524-762, 525-886, 532-1143, 541-856, 541-865, 541-1014, 541-1015, 541-1032, 541-1043, 541-1051, 541-1114, 541-1122, 542-822, 557-1043, 561-1239, 564-1247, 565-869, 565-1042, 573-767, 583-777, 583-784, 584-828, 585-861, 596-884, 596-928, 597-872, 598-710, 598-834, 610-1154, 640-901, 662-760, 664-1248, 668-968, 684-931, 698-963, 699-1168, 703-951, 705-969, 706-929, 706-958, 706-991, 707-926, 717-1220, 730-1020, 759-891, 772-1049, 773-991, 782-1047, 783-1081, 787-1248, 790-1054, 794-976, 795-1054, 795-1067, 795-1076, 813-1058, 813-1096, 831-1095, 869-1114, 869-1117, 875-1015, 875-1124, 897-1018, 921-1078, 935-1232, 935-1234, 935-1245, 971-1159, 985-1213, 987-1237, 1027-1248, 1041-1246, 1071-1202, 1109-1382, 1244-1393, 1246-1393, 1251-1393, 1261-1375, 1266-1393, 1279-1383, 1285-1393, 1287-1393, 1304-1382, 1309-1393
66/7509104CB1 1344	1-334, 1-353, 1-413, 1-441, 1-449, 1-457, 1-466, 1-482, 1-495, 1-515, 1-534, 1-546, 2-245, 3-460, 13-123, 13-212, 13-1336, 36-474, 50-878, 51-542, 312-521, 519-1289, 539-1155, 543-1096, 558-1281, 558-1289, 571-1289, 583-971, 655-1293, 676-1340, 705-978, 705-1004, 718-1254, 801-1322, 809-1302, 835-1131, 889-1318, 893-1344, 1061-1259, 1063-1309, 1069-1336, 1088-1340, 1164-1344

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
67/7509996CB1 2080	1-594, 10-304, 15-489, 15-505, 15-519, 15-556, 15-570, 15-643, 15-646, 15-647, 15-655, 15-660, 15-680, 15-693, 15-701, 17-430, 17-627, 18-545, 18-564, 18-604, 18-623, 18-627, 18-631, 18-670, 18-681, 27-297, 27-330, 27-336, 27-402, 27-424, 27-435, 27-540, 27-571, 27-578, 29-269, 30-204, 30-214, 30-261, 30-266, 30-270, 30-274, 30-286, 30-301, 30-310, 30-325, 30-327, 30-334, 30-339, 30-2080, 32-288, 32-567, 33-314, 33-322, 35-314, 35-655, 36-331, 36-339, 36-708, 37-288, 37-323, 37-369, 37-370, 37-571, 37-691, 39-344, 40-310, 40-319, 44-311, 44-345, 46-169, 46-273, 46-278, 46-290, 46-294, 46-303, 46-330, 46-333, 46-334, 46-335, 46-339, 46-342, 46-347, 46-357, 46-384, 46-535, 46-623, 46-627, 46-646, 46-653, 47-307, 49-204, 49-312, 49-591, 50-355, 50-436, 50-702, 51-357, 54-253, 54-704, 56-359, 56-439, 60-361, 61-497, 61-679, 62-277, 62-699, 66-859, 72-518, 74-747, 79-589, 84-378, 84-382, 87-672, 96-643, 107-699, 126-365, 131-780, 132-725, 136-707, 136-717, 142-417, 168-475, 168-635, 185-707, 191-770, 197-857, 210-541, 212-681, 215-800, 216-867, 219-674, 219-808, 220-865, 220-936, 223-772, 238-803, 240-463, 240-537, 242-476, 247-710, 248-827, 252-803, 255-503, 255-607, 259-907, 261-486, 270-596, 274-937, 281-516, 291-939, 303-773, 304-547, 304-576, 307-596, 308-508, 310-780, 310-817, 310-844, 314-767, 317-843, 318-542, 323-591, 323-905, 325-1007, 331-803, 337-643, 341-688, 342-747, 345-922, 351-1029, 352-994, 354-1020, 356-772, 358-707, 358-770, 368-571, 372-659, 375-735, 377-646, 382-627, 382-632, 387-1040, 389-904, 390-662, 399-874, 419-716, 427-874, 427-951, 428-1015, 432-1011, 437-688, 442-696, 443-762, 443-920, 448-1037, 449-905, 458-921, 458-980, 462-758, 493-874, 495-874, 500-718, 500-746, 538-844, 539-790, 542-770, 552-803, 564-787, 571-722, 582-814, 588-798, 594-834, 599-835, 608-835, 608-892, 628-925, 636-870, 679-913, 688-834, 688-929, 851-1419, 870-1420, 900-1419, 925-1410, 955-1175, 988-1041, 1050-1419, 1347-1917, 1347-1976, 1347-1987, 1347-2022, 1347-2026, 1347-2050, 1347-2073, 1350-1990, 1352-1958, 1355-1998, 1356-2075, 1357-2075, 1368-1608, 1369-1671, 1371-1651, 1375-1711, 1375-1879, 1375-1957, 1375-1962, 1375-1976, 1375-1977, 1375-2071, 1380-1523, 1383-1949, 1387-1738, 1387-1976, 1394-1863, 1397-1646, 1397-1648, 1398-1632, 1398-1635, 1398-1755, 1406-1647, 1408-1982, 1410-1923, 1424-1643, 1425-1645, 1426-2034, 1427-1718, 1431-2044, 1432-1599, 1432-1666, 1445-1986, 1449-1719, 1456-1976, 1462-1694, 1462-2079, 1471-1701, 1472-1720, 1484-2075, 1488-1989, 1489-1716, 1495-1999, 1498-2080, 1500-1905, 1500-2002, 1500-2062, 1501-2026, 1502-2030, 1510-1742, 1512-2030, 1513-1758, 1517-1736, 1524-1756, 1526-1793, 1526-1842, 1527-2066, 1532-2080, 1537-1837, 1537-2037, 1539-1763, 1542-1788, 1544-1814, 1544-1844, 1544-2070, 1545-1783, 1546-1766, 1551-1793, 1553-1764, 1553-1796, 1554-2030, 1560-1687, 1561-1844, 1573-1822, 1574-1816, 1574-1820, 1574-1884, 1580-1833, 1583-1719, 1587-2064, 1588-1843, 1589-1865, 1595-2032, 1596-1935, 1598-1773, 1600-2038, 1600-2078,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
67 - Cont'd	1601-2080, 1607-2080, 1632-1883, 1634-2080, 1636-2073, 1639-2074, 1641-1866, 1643-2076, 1644-1820, 1644-2074, 1648-1831, 1648-2076, 1654-1944, 1658-1747, 1663-1887, 1663-2080, 1665-2076, 1667-1982, 1671-1921, 1671-1975, 1672-2072, 1673-1899, 1673-1922, 1675-1885, 1675-1944, 1676-2074, 1680-1984, 1682-2078, 1687-1944, 1687-2075, 1687-2076, 1687-2080, 1689-1891, 1691-1930, 1696-2079, 1701-2018, 1712-1940, 1714-1949, 1715-1912, 1720-1973, 1727-1984, 1727-1991, 1748-2074, 1749-1996, 1751-2030, 1752-2048, 1754-1993, 1754-1999, 1754-2028, 1754-2062, 1764-2014, 1766-2075, 1768-1995, 1768-2008, 1768-2069, 1768-2070, 1768-2071, 1768-2073, 1768-2074, 1768-2075, 1768-2078, 1768-2079, 1768-2080, 1771-2030, 1772-1980, 1772-1992, 1773-1996, 1773-2000, 1774-2022, 1782-1989, 1782-2080, 1786-2076, 1803-1970, 1806-2076, 1818-2076, 1819-2080, 1821-1998, 1837-2041, 1841-2071, 1841-2072, 1841-2079, 1842-2076, 1844-2074, 1849-2080, 1852-2075, 1853-2072, 1855-2062, 1857-2072, 1859-2080, 1864-2080, 1866-2080, 1871-2080, 1873-2080, 1885-2069, 1885-2076, 1886-2062, 1896-2080, 1899-2063, 1918-2080, 1920-2080, 1930-2080, 1948-2080, 1953-2080, 1974-2080, 1976-2048, 1987-2080, 1988-2080, 1997-2080
68/7510030CB1 1418	1-1404, 710-916, 710-976, 938-1418
69/7510062CB1 3517	1-599, 1-3474, 65-719, 163-896, 163-951, 163-969, 163-1028, 163-1105, 163-1111, 165-1118, 166-1136, 169-937, 235-558, 245-679, 264-649, 285-710, 571-865, 584-1182, 657-840, 658-902, 658-1030, 658-1092, 658-1220, 658-1257, 658-1335, 658-1351, 691-932, 691-1254, 756-1252, 766-1330, 790-1305, 800-1362, 832-1177, 856-1119, 869-1440, 881-1152, 881-1346, 917-1429, 1015-1263, 1018-1046, 1018-1048, 1046-1242, 1059-1242, 1309-1430, 1336-1963, 1420-1645, 1438-1486, 1515-1764, 1516-1899, 1516-2158, 1516-2177, 1516-2201, 1516-2205, 1516-2214, 1517-2210, 1517-2214, 1531-1842, 1539-2145, 1551-2386, 1585-1929, 1586-2173, 1592-1897, 1603-2135, 1616-1902, 1688-1937, 1688-2214, 1688-2274, 1720-1937, 1763-2009, 1763-2242, 1773-2037, 1795-2300, 1831-2060, 1848-2472, 1863-2556, 1865-2155, 1876-2154, 1889-2622, 1913-2666, 1921-2419, 1959-2250, 1959-2268, 1959-2383, 1959-2486, 1959-2542, 1961-2188, 1961-2539, 1961-2545, 1967-2708, 1978-2639, 1987-2651, 1993-2376, 1996-2247, 2024-2782, 2027-2655, 2053-2265, 2057-2418,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
69 - Cont'd	2058-2479, 2058-2774, 2059-2287, 2060-2611, 2068-2679, 2077-2355, 2077-2732, 2095-2756, 2102-2359, 2102-2461, 2103-2321, 2121-2397, 2128-2701, 2177-2476, 2177-2587, 2178-2439, 2178-2804, 2182-2440, 2188-2478, 2191-2689, 2197-2601, 2200-2796, 2219-2486, 2219-2837, 2261-2910, 2302-2963, 2335-2596, 2339-2646, 2355-2960, 2360-3020, 2372-2630, 2408-3034, 2426-2790, 2437-3174, 2441-2745, 2461-2844, 2462-3073, 2487-2742, 2488-2729, 2489-2754, 2491-3102, 2514-2810, 2514-2953, 2546-2814, 2547-3152, 2548-2821, 2549-3030, 2562-3172, 2562-3173, 2598-2855, 2598-3177, 2611-2998, 2621-2823, 2621-3096, 2625-2968, 2626-2874, 2626-2884, 2626-3087, 2630-3256, 2648-3225, 2651-3233, 2659-3154, 2659-3215, 2660-3179, 2665-3073, 2675-3451, 2683-3453, 2685-2889, 2690-3101, 2700-3298, 2707-2907, 2707-2936, 2707-3000, 2707-3049, 2707-3189, 2707-3278, 2709-2944, 2714-3335, 2718-2977, 2718-3293, 2720-3002, 2720-3003, 2720-3370, 2722-2979, 2726-3285, 2727-3213, 2742-3354, 2742-3472, 2746-3221, 2746-3299, 2761-3248, 2767-3098, 2778-3018, 2778-3383, 2784-3349, 2786-2984, 2790-3455, 2791-3046, 2792-3198, 2793-3432, 2797-3414, 2801-3317, 2801-3429, 2801-3445, 2810-3410, 2812-3079, 2812-3343, 2815-3323, 2815-3402, 2816-3298, 2831-3457, 2837-3441, 2839-3147, 2848-3422, 2855-3182, 2856-3293, 2858-3314, 2858-3413, 2860-3274, 2876-3223, 2894-3470, 2905-3416, 2910-3402, 2911-3455, 2918-3450, 2920-3474, 2923-3101, 2929-3455, 2935-3244, 2943-3036, 2943-3480, 2950-3474, 2950-3479, 2952-3386, 2954-3228, 2954-3232, 2962-3454, 2965-3404, 2969-3469, 2976-3455, 2985-3455, 2988-3455, 2989-3455, 2995-3454, 2998-3454, 3000-3245, 3000-3279, 3000-3446, 3000-3455, 3006-3474, 3006-3476, 3007-3455, 3010-3455, 3022-3479, 3025-3469, 3026-3455, 3033-3455, 3035-3443, 3035-3455, 3035-3475, 3036-3455, 3046-3455, 3050-3455, 3051-3486, 3051-3517, 3054-3349, 3054-3455, 3055-3458, 3055-3459, 3059-3455, 3060-3455, 3064-3470, 3069-3454, 3077-3455, 3089-3455, 3093-3475, 3099-3256, 3100-3455, 3108-3455, 3110-3459, 3115-3457, 3129-3306, 3138-3455, 3147-3393, 3149-3353, 3161-3475, 3177-3455, 3182-3455, 3195-3298, 3195-3421, 3195-3443, 3217-3433, 3223-3455, 3237-3455, 3242-3455, 3276-3455, 3276-3517, 3280-3455, 3281-3455, 3285-3414, 3291-3455, 3342-3455, 3342-3470
70/7510217CB1 1472	1-142, 2-142, 38-1454, 45-801, 45-823, 45-949, 144-508, 145-368, 153-575, 153-578, 153-606, 154-538, 154-582, 164-571, 174-531, 200-365, 200-388, 200-534, 200-539, 200-582, 226-462, 228-510, 233-367, 272-549, 334-636, 373-585, 616-1106, 656-1100, 684-926, 737-1271, 743-1274, 793-1274, 799-1085, 799-1258, 831-1274, 833-1077, 851-1111, 851-1146, 851-1271, 851-1274, 874-1187, 894-1274, 901-1189, 999-1274, 1007-1406, 1012-1274, 1038-1460, 1052-1471, 1054-1460, 1065-1459, 1066-1454, 1084-1452, 1096-1435, 1147-1356, 1147-1457, 1171-1471, 1182-1458, 1182-1472

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
71/7510298CB1 2160	1-619, 1-637, 1-2144, 7-525, 15-217, 20-361, 27-241, 269-848, 290-687, 324-556, 342-707, 548-725, 804-906, 804-1167, 804-1203, 806-1058, 806-1438, 807-1248, 812-1285, 818-1255, 854-1465, 856-1114, 866-1493, 875-1383, 886-1202, 888-1628, 909-1499, 914-1524, 926-1188, 930-1497, 938-1523, 951-1577, 956-1497, 961-1556, 973-1238, 979-1161, 989-1540, 999-1609, 1014-1556, 1014-1618, 1029-1336, 1031-1201, 1045-1610, 1048-1300, 1048-1632, 1053-1319, 1059-1196, 1073-1715, 1092-1691, 1094-1388, 1097-1499, 1099-1694, 1101-1686, 1105-1723, 1111-1739, 1121-1685, 1132-1361, 1136-1679, 1137-1388, 1143-1767, 1151-1689, 1159-1728, 1173-1501, 1186-1758, 1186-1781, 1195-1380, 1195-1409, 1197-1714, 1198-1470, 1200-1836, 1203-1321, 1204-1493, 1207-1508, 1209-1779, 1217-1874, 1218-1813, 1222-1313, 1228-1733, 1228-2136, 1231-1339, 1242-2083, 1245-1575, 1246-1482, 1247-1347, 1248-1497, 1256-1652, 1258-1494, 1259-1550, 1259-1568, 1259-1579, 1261-2078, 1266-1954, 1267-1518, 1277-1822, 1284-2135, 1287-1579, 1292-1841, 1293-2107, 1294-1845, 1294-1854, 1294-1885, 1294-1896, 1294-1939, 1303-1531, 1304-2136, 1309-1581, 1310-1536, 1325-1778, 1328-2136, 1329-1446, 1329-1748, 1332-1609, 1332-1614, 1335-1545, 1335-1562, 1335-1581, 1335-1890, 1336-1547, 1336-2136, 1338-1602, 1339-1508, 1345-1628, 1345-2156, 1353-2160, 1354-2135, 1354-2136, 1359-2136, 1362-1696, 1365-1612, 1365-1978, 1370-2154, 1370-2159, 1372-1608, 1378-1591, 1379-1717, 1379-2074, 1385-1831, 1402-2135, 1407-1776, 1407-1782, 1407-2136, 1411-2136, 1417-1958, 1428-1682, 1429-2142, 1430-1677, 1433-1615, 1433-2160, 1435-1673, 1435-2146, 1437-2074, 1440-2160, 1442-2135, 1446-1627, 1446-1724, 1446-1727, 1450-1837, 1452-2010, 1456-1961, 1456-2098, 1459-2057, 1462-1866, 1467-2070, 1479-2160, 1500-1803, 1501-2089, 1502-1764, 1505-1822, 1513-2159, 1519-2060, 1520-1790, 1523-2144, 1529-2136, 1530-1786, 1535-1852, 1544-2133, 1547-1785, 1548-1811, 1549-2159, 1552-1778, 1558-2043, 1560-1843, 1560-2061, 1563-2136, 1564-1845, 1574-2160, 1577-1782, 1577-2160, 1581-2160, 1589-2160, 1591-2043, 1598-2136, 1599-2150, 1602-1845, 1606-1958, 1606-2150, 1611-2160, 1614-2035, 1629-2160, 1633-2135, 1636-2125, 1639-2116, 1647-2108, 1656-2160, 1658-2146, 1660-1932, 1661-1917, 1664-1906, 1664-2160, 1666-2160, 1667-2136, 1670-2136, 1673-2146, 1681-2143, 1683-2146, 1690-2145, 1695-2116, 1700-1985, 1700-2160, 1703-2152, 1712-2154, 1715-2160, 1716-2100, 1720-2160, 1727-2160, 1732-2146, 1737-2146, 1737-2155, 1740-2155, 1741-2146, 1742-2153, 1743-2146, 1744-2146, 1747-2015, 1760-2013, 1760-2027, 1768-2146, 1778-2160, 1779-2146, 1780-2060, 1780-2146, 1785-2144, 1786-2148, 1789-2147, 1803-2146, 1804-2146, 1805-2146, 1806-2147, 1809-2107, 1812-2151, 1820-2144, 1828-2009, 1828-2046, 1828-2117, 1828-2160, 1831-2090, 1832-2090, 1833-2142, 1837-2160, 1868-2146, 1868-2159, 1871-2146, 1887-2146, 1891-2146, 1892-2151, 1892-2153, 1897-2148, 1899-2146, 1906-2160, 1907-2146, 1926-2160, 1927-2146, 1929-2160, 1931-2160, 1940-2155, 1949-2160, 1956-2159, 1971-2146, 2000-2105, 2002-2160, 2005-2146, 2005-2160, 2006-2160, 2008-2152, 2058-2149, 2058-2160

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
72/7510299CB1 2383	1-268, 1-346, 2-274, 4-245, 10-2383, 14-298, 20-263, 20-268, 23-732, 24-708, 51-672, 51-751, 77-673, 103-790, 137-674, 147-405, 157-590, 157-619, 157-622, 157-644, 157-649, 157-656, 157-658, 157-716, 157-731, 157-737, 157-741, 157-869, 157-883, 157-895, 157-896, 157-926, 157-930, 158-833, 163-383, 165-821, 170-294, 170-397, 170-516, 170-606, 170-611, 170-720, 170-818, 170-870, 170-889, 170-968, 170-1079, 173-699, 174-875, 189-968, 191-631, 191-632, 198-879, 217-846, 224-986, 224-1041, 267-502, 267-611, 267-700, 267-787, 267-788, 267-807, 270-883, 297-1067, 301-542, 306-851, 307-906, 320-1020, 326-952, 332-849, 337-1024, 346-445, 359-910, 369-1071, 385-978, 390-970, 410-1052, 422-1043, 462-750, 462-768, 477-584, 488-735, 500-774, 506-742, 533-784, 533-822, 533-876, 553-808, 554-784, 566-752, 576-814, 577-787, 582-880, 584-781, 606-805, 636-908, 640-889, 672-777, 698-1016, 711-992, 784-982, 880-997, 1040-1433, 1040-1591, 1041-1302, 1061-1265, 1270-1481, 1290-1936, 1308-1855, 1308-1871, 1334-2164, 1344-1563, 1348-1459, 1354-1814, 1384-1638, 1384-1818, 1390-1910, 1395-1595, 1398-1612, 1398-1686, 1402-1589, 1404-1658, 1406-2027, 1406-2243, 1409-2001, 1433-1707, 1434-1954, 1439-2005, 1467-1573, 1468-1756, 1488-2190, 1488-2315, 1494-1897, 1495-1976, 1504-2184, 1506-2133, 1508-1935, 1513-1751, 1514-1745, 1527-1829, 1528-2306, 1533-2040, 1533-2223, 1534-1871, 1534-1951, 1535-2093, 1538-2259, 1543-2267, 1544-2294, 1582-1837, 1584-2261, 1586-1846, 1592-2326, 1597-2306, 1599-2226, 1614-1868, 1614-1871, 1623-1858, 1623-2303, 1624-2328, 1631-1832, 1637-2261, 1640-2235, 1654-2295, 1681-2321, 1682-2328, 1689-1901, 1696-1855, 1703-2029, 1713-2069, 1722-2324, 1727-1888, 1734-2208, 1736-1962, 1736-2017, 1742-2274, 1756-2328, 1761-2328, 1767-2018, 1767-2258, 1767-2315, 1776-2014, 1776-2226, 1776-2253, 1777-2038, 1779-2027, 1785-2048, 1788-2328, 1789-2195, 1790-1960, 1795-2328, 1796-2328, 1807-2328, 1812-2328, 1819-2137, 1825-2328, 1835-2359, 1835-2373, 1836-2096, 1836-2229, 1836-2328, 1843-2328, 1847-2325, 1849-2328, 1860-2296, 1861-2328, 1865-2076, 1865-2130, 1865-2260, 1865-2328, 1866-2328, 1868-2328, 1872-2327, 1872-2328, 1873-2302, 1873-2328, 1874-2328, 1875-2325, 1875-2338, 1876-2023, 1877-2328, 1878-2109, 1878-2231, 1878-2328, 1879-2323, 1879-2327, 1879-2328, 1880-2327, 1880-2328, 1881-2328, 1882-2325, 1882-2328, 1883-2328, 1884-2325, 1884-2328, 1885-2323, 1886-2328, 1888-2326, 1888-2328, 1890-2150, 1891-2328, 1895-2112, 1895-2328, 1898-2094, 1898-2328, 1901-2343, 1902-2325, 1903-2328, 1905-1999, 1906-2328, 1907-2328, 1908-2328, 1913-2325, 1914-2328, 1918-2325, 1918-2328, 1919-2328, 1920-2328, 1920-2338, 1921-2130, 1921-2325, 1921-2328, 1925-2328, 1931-2328, 1934-2328, 1937-2328, 1939-2325, 1939-2327, 1939-2328, 1940-2328, 1941-2325, 1949-2327, 1950-2328, 1952-2325, 1952-2328, 1953-2220, 1955-2328, 1956-2328, 1962-2298, 1965-2328, 1967-2328, 1969-2328, 1979-2328, 1981-2244, 1982-2328, 1983-2328, 1986-2328, 1988-2325, 1990-2328, 1991-2325, 1991-2328, 1991-2373, 1996-2328,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
72 - Cont'd	2010-2328, 2013-2328, 2014-2324, 2017-2328, 2030-2328, 2039-2297, 2049-2142, 2054-2328, 2060-2328, 2072-2325, 2073-2298, 2075-2201, 2079-2328, 2082-2328, 2089-2325, 2103-2327, 2107-2328, 2111-2328, 2112-2340, 2113-2325, 2119-2324, 2121-2325, 2123-2325, 2123-2328, 2125-2323, 2125-2328, 2125-2368, 2126-2275, 2126-2325, 2126-2328, 2127-2297, 2127-2328, 2129-2328, 2130-2328, 2131-2321, 2133-2328, 2134-2328, 2136-2328, 2141-2327, 2145-2328, 2149-2338, 2151-2274, 2154-2328, 2157-2328, 2162-2328, 2163-2328, 2165-2383, 2185-2328, 2190-2373, 2196-2328, 2210-2287, 2210-2373, 2222-2328, 2244-2328, 2248-2369, 2262-2328, 2263-2373
73/7510368CB1 1819	1-221, 1-233, 1-276, 1-409, 1-726, 1-1809, 13-189, 13-291, 14-297, 16-279, 24-499, 31-300, 42-331, 63-291, 93-290, 162-367, 194-773, 203-675, 206-771, 206-776, 209-760, 293-776, 479-709, 506-777, 537-796, 612-937, 673-911, 716-1288, 736-1266, 741-999, 779-1075, 783-1066, 793-1104, 794-1073, 795-1217, 805-1265, 831-972, 868-1130, 874-1431, 886-1141, 888-1043, 900-1113, 923-1557, 923-1798, 959-1241, 965-1798, 966-1665, 978-1464, 986-1584, 996-1631, 1009-1407, 1013-1265, 1021-1462, 1027-1305, 1027-1584, 1048-1753, 1062-1797, 1063-1610, 1072-1798, 1085-1627, 1085-1757, 1100-1691, 1118-1357, 1119-1668, 1123-1498, 1129-1380, 1130-1798, 1166-1433, 1176-1411, 1176-1426, 1189-1310, 1212-1798, 1220-1798, 1231-1661, 1239-1494, 1246-1495, 1247-1760, 1251-1808, 1265-1715, 1266-1752, 1271-1496, 1277-1536, 1299-1798, 1305-1579, 1309-1509, 1309-1518, 1316-1546, 1328-1819, 1333-1798, 1341-1795, 1342-1798, 1349-1790, 1365-1789, 1375-1819, 1379-1798, 1383-1798, 1396-1798, 1401-1819, 1418-1759, 1464-1722, 1468-1619, 1478-1788, 1481-1808, 1486-1798, 1494-1798, 1495-1798, 1496-1798, 1497-1805, 1512-1806, 1516-1742, 1532-1806, 1540-1794, 1544-1796, 1579-1798, 1589-1798, 1600-1798, 1608-1819, 1618-1810, 1619-1808, 1662-1798, 1685-1819, 1733-1798
74/7510369CB1 1796	1-397, 1-419, 1-544, 17-249, 17-425, 17-1789, 29-205, 30-313, 32-295, 79-307, 109-306, 138-264, 188-425, 213-510, 213-513, 225-654, 246-832, 254-513, 304-552, 309-426, 313-641, 403-863, 424-763, 424-767, 430-915, 460-887, 486-729, 492-817, 553-791, 616-1146, 621-879, 659-955, 674-953, 675-1097, 685-1145, 709-1145, 748-1010, 766-1021, 780-993, 890-1287, 1185-1453, 1191-1415, 1191-1571, 1226-1786, 1226-1787, 1267-1508, 1274-1779, 1320-1790, 1322-1733, 1401-1442, 1401-1478, 1401-1570, 1401-1607, 1401-1732, 1401-1736, 1401-1777, 1401-1778, 1401-1780, 1401-1784, 1401-1788, 1401-1796, 1403-1777, 1444-1702, 1448-1599, 1461-1786, 1466-1776, 1474-1776, 1475-1776, 1476-1776, 1477-1785, 1492-1786, 1512-1786, 1520-1774, 1559-1776, 1569-1776, 1580-1776, 1642-1776, 1713-1777

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
75/7510377CB1 2635	1-2516, 296-347, 1056-1326, 1090-1907, 1113-1366, 1255-1512, 1289-1934, 1326-1584, 1339-1858, 1348-1605, 1348-1612, 1348-1825, 1348-1834, 1348-1854, 1348-1858, 1348-1865, 1348-1867, 1360-2028, 1396-1992, 1447-1697, 1447-1698, 1447-1795, 1447-1831, 1447-1906, 1447-1962, 1471-1975, 1475-2078, 1480-2082, 1505-1771, 1505-1775, 1506-1973, 1509-1768, 1525-2037, 1527-2094, 1533-2069, 1537-2078, 1543-1823, 1543-2099, 1566-2097, 1567-2114, 1584-1808, 1588-2043, 1610-1868, 1615-2466, 1619-1841, 1626-1838, 1627-2320, 1629-1887, 1629-2022, 1629-2060, 1630-2099, 1656-1932, 1662-2178, 1664-2287, 1674-2218, 1677-2393, 1679-2232, 1684-2323, 1685-1949, 1705-1877, 1719-2387, 1723-2163, 1724-1813, 1776-2305, 1791-1845, 1805-2080, 1807-2316, 1810-2118, 1810-2297, 1813-2061, 1827-2107, 1827-2114, 1827-2131, 1828-1972, 1836-2203, 1838-2432, 1845-2072, 1887-2087, 1890-2166, 1894-2491, 1895-2491, 1896-2437, 1898-2446, 1902-2449, 1904-2377, 1916-2411, 1939-2452, 1942-2163, 1943-2486, 1949-2488, 1953-2156, 1967-2470, 1969-2386, 1969-2447, 1974-2237, 1985-2448, 1995-2398, 2003-2445, 2004-2416, 2007-2280, 2024-2479, 2027-2488, 2027-2490, 2031-2488, 2035-2491, 2041-2635, 2045-2490, 2062-2488, 2074-2459, 2077-2513, 2078-2488, 2084-2534, 2087-2491, 2088-2488, 2101-2488, 2101-2534, 2109-2492, 2131-2484, 2141-2490, 2144-2511, 2150-2450, 2152-2449, 2188-2518, 2190-2296, 2192-2511, 2197-2472, 2207-2487, 2214-2475, 2221-2490, 2236-2407, 2260-2491, 2278-2524, 2313-2491, 2401-2491, 2450-2516
76/7510026CB1 4046	1-3874, 193-1044, 206-1044, 209-856, 216-1044, 343-931, 354-988, 389-1078, 473-1017, 480-1018, 485-980, 485-1027, 534-1205, 556-1025, 649-1188, 664-1251, 734-1052, 766-1054, 806-1553, 885-1256, 1037-1475, 1037-1477, 1303-1537, 1303-1666, 1305-1931, 1355-1758, 1356-1941, 1394-1597, 1420-2130, 1450-2354, 1497-2154, 1526-2127, 1537-1986, 1537-2105, 1538-1969, 1538-2235, 1607-2221, 1622-2237, 1675-2205, 1679-1932, 1690-2305, 1711-2323, 1712-2172, 1712-2340, 1784-2367, 1788-2462, 1796-2342, 1814-2461, 1824-2142, 1824-2570, 1846-2525, 1853-2637, 1866-2641, 1870-2440, 1872-2682, 1889-2487, 1889-2607, 1902-2569, 1916-2555, 1929-2529, 1932-2580, 1933-2496, 1984-2580, 1996-2872, 2011-2704, 2017-2500, 2028-2306, 2048-2670, 2053-2280, 2072-2322, 2077-2484, 2085-2663, 2100-2597, 2125-2650, 2129-2412, 2157-2689, 2171-2570, 2172-2895, 2184-2312, 2189-2814, 2191-2488, 2192-3018, 2231-3004, 2242-2511, 2248-2589, 2259-2789, 2265-3019, 2285-2906, 2348-2957, 2365-2755, 2372-2672, 2375-2984, 2378-3276, 2416-2673,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
76 - Cont'd	2430-3277, 2481-3009, 2496-2717, 2499-2709, 2499-3049, 2506-3123, 2509-3180, 2524-3069, 2539-3197, 2595-2870, 2605-3343, 2608-3271, 2626-3241, 2639-3200, 2647-2851, 2652-2921, 2655-2728, 2665-2728, 2677-3197, 2680-3137, 2695-3281, 2710-3399, 2730-2993, 2763-3030, 2765-3361, 2774-2896, 2800-3132, 2835-3519, 2842-3180, 2849-3388, 2877-3363, 2877-3527, 2878-3560, 2879-3382, 2888-3518, 2902-3451, 2902-3560, 2911-3149, 2916-3464, 2919-3436, 2932-3459, 2938-3629, 2942-3438, 2953-3238, 2954-3765, 2955-3462, 2955-3513, 2962-3500, 2982-3583, 2988-3445, 2994-3542, 2997-3241, 3005-3263, 3014-3557, 3018-3293, 3022-3443, 3030-3497, 3091-3713, 3099-3483, 3133-3641, 3134-3713, 3135-3727, 3143-3587, 3147-3623, 3160-3623, 3160-3758, 3162-3568, 3169-3417, 3176-3769, 3180-3623, 3186-3478, 3190-3641, 3192-3643, 3193-3643, 3201-3623, 3205-3634, 3213-3624, 3225-3515, 3235-3827, 3261-3820, 3269-3624, 3281-3664, 3282-3805, 3288-3482, 3292-3524, 3293-3643, 3298-3874, 3299-3564, 3333-3864, 3347-3667, 3353-3721, 3395-3662, 3402-3624, 3427-3677, 3458-4046, 3474-3699, 3480-3626, 3482-3778, 3513-3845, 3550-3806, 3560-3816
77/7509168CB1 858	1-840, 2-238, 4-325, 11-238, 52-780, 196-433, 240-577, 309-558, 330-568, 433-756, 442-706, 442-858, 469-837, 503-843, 557-848, 570-850, 661-836, 661-858, 728-856, 728-858
78/7500607CB1 3129	1-273, 1-521, 14-296, 17-521, 30-291, 30-292, 30-651, 30-3129, 33-267, 37-3126, 38-223, 60-331, 64-310, 73-263, 82-318, 83-314, 128-339, 153-682, 285-528, 285-705, 314-715, 314-717, 378-659, 395-956, 402-668, 412-730, 421-719, 451-715, 451-720, 508-651, 530-1087, 552-749, 556-1183, 580-836, 583-855, 652-1259, 663-1328, 719-1281, 725-1324, 758-1333, 780-1388, 783-1311, 793-1398, 801-1319, 818-1109, 899-1149, 899-1468, 899-1527, 899-1549, 899-1682, 907-1007, 908-1464, 915-1460, 932-1225, 942-1583, 944-1605, 976-1275, 976-1298, 1006-1536, 1007-1247, 1007-1376, 1011-1494, 1039-1355, 1060-1638, 1101-1694, 1109-1319, 1110-1527, 1142-1783, 1146-1370, 1159-1720, 1166-1728, 1169-1668, 1175-1328, 1198-1447, 1228-1723, 1229-1807, 1238-1779, 1252-1855, 1253-1704, 1253-1830, 1285-1834, 1287-1834, 1299-1847, 1301-1883, 1307-1570, 1309-1516, 1310-1846, 1310-1850, 1345-2060, 1354-1908, 1364-1667, 1393-1646, 1397-1974, 1415-1895, 1444-2022, 1447-1628, 1469-1954, 1475-2003, 1526-2016, 1565-2386, 1605-1896,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
78 - Cont'd	1665-2369, 1755-2059, 1755-2143, 1755-2171, 1755-2205, 1755-2217, 1755-2231, 1755-2238, 1755-2270, 1755-2296, 1755-2316, 1755-2327, 1755-2330, 1755-2335, 1755-2338, 1755-2340, 1755-2343, 1755-2344, 1755-2366, 1755-2371, 1755-2387, 1756-2000, 1762-1956, 1779-2034, 1779-2207, 1782-2658, 1798-1968, 1806-2301, 1823-2084, 1856-2154, 1864-2668, 1916-2185, 1945-2379, 1948-2219, 1949-2201, 1951-2575, 1960-2216, 1963-2221, 1997-2680, 2000-2483, 2025-2670, 2026-2630, 2027-2303, 2029-2303, 2033-2602, 2039-2400, 2046-2243, 2046-2589, 2047-2285, 2047-2519, 2051-2328, 2063-2618, 2066-2586, 2080-2377, 2080-2652, 2082-2703, 2095-2714, 2099-2343, 2109-2670, 2109-2742, 2114-2385, 2116-2299, 2117-2692, 2129-2339, 2135-2644, 2141-2315, 2151-2433, 2153-2833, 2158-2337, 2160-2292, 2160-2610, 2163-2340, 2163-2662, 2169-2463, 2169-2679, 2170-2662, 2179-2312, 2184-2706, 2195-2497, 2208-2591, 2209-2591, 2218-2331, 2241-2440, 2243-2468, 2251-2416, 2288-2562, 2289-2599, 2302-2792, 2302-2796, 2302-2858, 2302-2869, 2302-2915, 2302-2937, 2303-2613, 2303-2761, 2303-2796, 2303-2817, 2303-2848, 2303-2871, 2303-2876, 2303-2918, 2303-2935, 2303-2937, 2303-2946, 2311-2595, 2321-2606, 2330-2726, 2337-2899, 2353-2593, 2410-2741, 2417-2955, 2424-2705, 2435-3009, 2439-2688, 2446-2883, 2456-2701, 2469-2661, 2471-2669, 2471-2748, 2472-2987, 2492-3129, 2646-2882, 2663-2955, 3013-3086
79/7506079CB1 1520	1-91, 1-1508, 8-618, 8-788, 51-130, 153-374, 153-431, 165-418, 171-754, 180-685, 187-459, 187-515, 200-435, 200-458, 200-468, 201-771, 201-859, 232-860, 250-782, 256-530, 323-534, 328-857, 333-637, 333-653, 338-532, 346-573, 349-618, 364-562, 366-625, 382-619, 414-988, 421-698, 435-932, 460-1118, 466-725, 467-880, 505-743, 507-1127, 520-796, 525-1287, 526-779, 545-1210, 551-1210, 557-1124, 559-748, 564-820, 572-813, 572-1210, 584-816, 607-876, 608-1502, 610-1164, 624-1502, 627-853, 634-1502, 639-833, 639-912, 646-925, 647-1502, 650-833, 652-1234, 659-824, 661-868, 662-1223, 671-1154, 671-1222, 675-1155, 679-1500, 680-1155, 682-1099, 685-1502, 687-1502, 696-1279, 708-1502, 718-1502, 719-1302, 728-988, 730-920, 759-1258, 773-1274, 783-1499, 790-1502, 794-1296, 801-1503, 813-1366, 813-1502, 824-1502, 824-1513, 825-1502, 835-1063, 835-1081, 835-1210, 838-1502, 847-1088, 850-1464, 885-1134, 886-1131, 898-1502, 913-1176, 923-1163, 926-1504, 927-1191, 928-1465, 952-1205, 957-1314, 960-1219, 970-1491,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
79 - Cont'd	995-1244, 995-1502, 1004-1290, 1004-1491, 1006-1469, 1008-1493, 1024-1508, 1025-1466, 1026-1520, 1034-1313, 1034-1507, 1035-1178, 1035-1278, 1036-1302, 1038-1266, 1038-1315, 1038-1328, 1038-1337, 1040-1204, 1048-1313, 1048-1356, 1048-1499, 1054-1502, 1061-1508, 1066-1202, 1075-1508, 1096-1378, 1100-1520, 1101-1520, 1109-1508, 1115-1507, 1129-1502, 1136-1447, 1138-1508, 1145-1507, 1155-1345, 1161-1508, 1172-1469, 1185-1502, 1186-1502, 1189-1485, 1198-1508, 1206-1428, 1211-1471, 1211-1508, 1215-1508, 1222-1398, 1226-1502, 1230-1502, 1230-1508, 1245-1508, 1286-1508, 1301-1507, 1307-1520, 1315-1508, 1318-1520, 1319-1505, 1339-1516, 1345-1508, 1356-1508, 1357-1507, 1399-1519, 1399-1520, 1400-1495, 1419-1507, 1457-1508
80/7509259CB1 575	1-148, 1-575, 221-575, 286-507, 287-539, 354-572, 488-575
81/7509263CB1 1635	1-424, 1-1635, 132-370, 150-1011, 153-1022, 196-457, 196-493, 197-311, 197-359, 197-377, 197-396, 197-408, 197-417, 197-428, 197-429, 197-432, 197-433, 197-434, 197-436, 197-438, 197-441, 197-444, 197-445, 197-446, 197-447, 197-448, 197-449, 197-450, 197-453, 197-456, 197-459, 197-466, 197-470, 197-472, 197-488, 197-492, 197-493, 198-424, 198-425, 198-450, 198-469, 199-442, 199-446, 199-450, 199-451, 199-457, 200-395, 200-436, 200-439, 200-441, 200-442, 200-451, 200-465, 200-467, 200-486, 202-446, 202-467, 203-442, 203-474, 204-480, 206-325, 206-345, 206-391, 206-414, 206-479, 206-490, 207-403, 207-493, 208-440, 217-493, 294-493, 431-1018, 494-644, 494-666, 494-690, 494-695, 494-703, 494-724, 494-734, 494-763, 494-773, 494-774, 494-840, 494-885, 494-953, 494-993, 494-1007, 494-1101, 502-1040, 503-795, 504-1040, 510-753, 510-777, 510-1104, 515-711, 515-1047, 516-1011, 517-1074, 522-770, 528-758, 529-1138, 533-623, 533-782, 534-774, 534-973, 562-794, 562-823, 567-1232, 568-1196, 572-851, 572-865, 576-848, 582-1140, 585-1112, 592-816, 594-870, 594-928, 594-1117, 619-645, 625-805, 625-957, 632-880, 637-844, 643-865, 646-955, 653-1122, 654-1218, 658-889, 658-1081, 659-1182, 659-1318, 662-963, 663-869, 665-1318, 669-940, 673-912, 675-909, 681-1076, 682-949, 683-891, 686-1051, 688-915, 693-952, 695-1173, 698-1103, 705-922, 705-993, 710-1166, 715-941, 716-1105, 719-972, 721-938, 740-1455, 742-1355, 744-1313, 746-1274, 748-1370, 749-1184, 751-1338, 752-1366, 757-1363, 762-1052, 763-1296, 765-1141, 767-988, 767-1103, 772-1074, 773-1074, 775-1012, 796-1042, 796-1068, 798-1353, 808-1085, 809-1390, 810-1263, 811-1024, 826-1635, 835-1361, 837-1327, 838-1131, 851-1103, 861-1554, 868-1020, 873-1166, 873-1323, 873-1326, 879-1126, 888-1633, 899-1296, 899-1301, 900-1549, 904-1611, 905-1159, 917-1350, 918-1105, 921-1159, 923-1140, 923-1163, 925-1556, 934-1123, 936-1578, 937-1286, 938-1162, 938-1174, 942-1202, 949-1138, 954-1123, 964-1209, 967-1626, 971-1635, 973-1209, 973-1264, 973-1454,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
81 - Cont'd	979-1635, 983-1446, 984-1110, 984-1218, 991-1247, 993-1635, 997-1366, 999-1252, 1003-1209, 1005-1238, 1007-1589, 1010-1202, 1011-1366, 1027-1314, 1029-1289, 1035-1590, 1042-1553, 1043-1366, 1059-1293, 1062-1305, 1066-1287, 1069-1527, 1071-1340, 1073-1628, 1078-1198, 1082-1301, 1083-1322, 1086-1313, 1086-1324, 1086-1625, 1087-1320, 1089-1345, 1093-1628, 1095-1628, 1109-1635, 1121-1358, 1127-1340, 1133-1359, 1146-1398, 1150-1628, 1154-1635, 1157-1590, 1158-1416, 1160-1409, 1160-1442, 1160-1445, 1168-1631, 1169-1372, 1169-1627, 1169-1631, 1169-1635, 1170-1628, 1170-1635, 1172-1599, 1172-1616, 1172-1631, 1173-1628, 1174-1407, 1174-1628, 1177-1632, 1178-1457, 1179-1416, 1179-1447, 1179-1628, 1180-1627, 1183-1439, 1192-1417, 1192-1454, 1193-1463, 1194-1363, 1194-1628, 1196-1628, 1198-1628, 1200-1635, 1202-1629, 1206-1628, 1207-1632, 1210-1635, 1214-1630, 1216-1635, 1219-1635, 1223-1628, 1224-1419, 1227-1628, 1227-1630, 1230-1480, 1230-1635, 1233-1482, 1234-1634, 1235-1628, 1237-1628, 1238-1628, 1239-1632, 1241-1628, 1242-1514, 1242-1633, 1242-1634, 1243-1626, 1245-1628, 1247-1632, 1247-1634, 1249-1568, 1249-1628, 1249-1629, 1261-1599, 1266-1516, 1274-1531, 1274-1626, 1278-1628, 1290-1630, 1303-1529, 1304-1632, 1311-1538, 1312-1628, 1322-1552, 1330-1549, 1331-1602, 1332-1505, 1344-1628, 1346-1487, 1353-1604, 1367-1629, 1367-1630, 1368-1628, 1368-1635, 1371-1634, 1387-1635, 1389-1525, 1398-1629, 1403-1626, 1404-1635, 1406-1628, 1412-1628, 1425-1628, 1427-1626, 1431-1628, 1438-1629, 1440-1634, 1441-1633, 1442-1616, 1464-1611, 1465-1628, 1507-1628, 1510-1626, 1513-1633, 1529-1635, 1548-1635
82/7509360CB1 1638	1-196, 9-318, 12-542, 24-286, 25-242, 25-267, 25-278, 25-290, 25-323, 25-330, 25-411, 25-439, 25-556, 25-587, 25-1497, 28-266, 28-287, 28-302, 28-616, 30-216, 30-279, 30-292, 30-330, 30-459, 30-500, 30-509, 30-555, 30-559, 30-579, 30-617, 31-580, 32-256, 33-171, 33-294, 33-295, 33-465, 33-893, 34-227, 34-593, 35-151, 35-223, 38-290, 39-282, 40-545, 42-571, 56-310, 56-324, 74-272, 85-337, 86-343, 349-687, 482-1371, 698-956, 698-1177, 698-1193, 698-1200, 698-1218, 698-1342, 699-1371, 700-1371, 701-1306, 701-1371, 704-1480, 715-1481, 729-1356, 732-1205, 736-1371, 739-1445, 742-1331, 743-1010, 750-831, 750-1210, 750-1371, 760-1296, 763-1210, 764-987, 767-1100, 767-1405, 767-1410, 767-1435, 773-1058, 773-1453, 798-1502, 803-1496, 804-1494, 806-978, 806-1049, 807-1452, 816-1496, 818-1052, 818-1394, 823-1002, 826-1207, 829-1071, 829-1499, 832-1431, 835-1442, 836-1471, 838-1476, 843-1448, 843-1495, 844-1073, 867-1498, 872-1300, 873-1494, 879-1476, 886-1236, 888-1471, 888-1503, 897-1137, 901-1481,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
82 - Cont'd	905-1447, 906-1515, 915-1172, 917-1119, 917-1157, 918-1216, 926-1074, 928-1454, 929-1111, 934-1150, 935-1497, 937-1179, 942-1442, 943-1161, 949-1492, 958-1498, 959-1445, 969-1496, 982-1481, 983-1444, 989-1370, 990-1324, 993-1315, 1000-1481, 1001-1478, 1010-1504, 1014-1504, 1021-1251, 1026-1307, 1026-1499, 1030-1335, 1031-1512, 1033-1478, 1037-1481, 1046-1211, 1048-1481, 1049-1245, 1049-1290, 1054-1479, 1054-1481, 1056-1487, 1057-1505, 1058-1481, 1059-1478, 1059-1481, 1060-1481, 1064-1481, 1096-1481, 1097-1443, 1105-1330, 1111-1444, 1120-1481, 1121-1427, 1125-1481, 1128-1481, 1129-1481, 1131-1340, 1133-1481, 1134-1481, 1135-1498, 1143-1481, 1152-1396, 1172-1473, 1173-1413, 1215-1520, 1219-1453, 1224-1487, 1226-1444, 1230-1512, 1263-1481, 1279-1495, 1287-1526, 1299-1506, 1304-1487, 1313-1484, 1313-1527, 1334-1638
83/7509394CB1 2354	1-124, 1-168, 1-209, 1-225, 1-227, 1-238, 1-239, 1-247, 1-254, 1-255, 1-263, 1-275, 1-283, 1-288, 1-324, 1-406, 1-412, 1-413, 1-433, 1-489, 1-502, 1-525, 1-531, 1-2322, 2-127, 2-464, 2-531, 3-263, 3-443, 7-265, 7-272, 8-265, 10-247, 10-269, 10-493, 10-619, 10-763, 11-531, 16-490, 16-526, 18-493, 21-203, 24-531, 26-396, 26-530, 28-276, 29-464, 29-493, 33-501, 63-531, 90-442, 105-297, 105-362, 106-407, 110-366, 154-401, 159-387, 200-457, 207-478, 272-504, 402-526, 530-741, 531-809, 540-1012, 558-810, 558-1128, 559-1065, 569-1166, 576-1138, 586-1452, 606-1137, 617-901, 617-1232, 619-877, 638-1063, 659-1184, 659-1211, 662-1143, 663-947, 672-1041, 681-1351, 687-1218, 689-918, 704-1364, 706-1207, 709-934, 711-1260, 714-879, 714-1296, 719-1568, 733-987, 739-1364, 741-1275, 750-775, 750-1524, 752-1557, 754-1262, 760-1386, 770-1348, 774-1141, 774-1191, 780-1031, 782-1325, 785-1274, 786-1254, 787-1046, 788-1230, 797-1310, 805-1350, 805-1438, 808-1369, 813-1061, 815-1078, 819-1243, 822-1260, 828-1105, 833-1481, 840-1577, 856-1298, 857-1570, 863-1529, 864-1529, 868-1137, 869-1191, 869-1260, 869-1497, 870-1490, 874-1666, 888-1160, 888-1373, 895-1140, 900-1676, 904-1141, 906-1393, 908-1388, 908-1456, 910-1179, 911-1155, 911-1177, 919-1207, 919-1214, 931-1288, 941-1469, 949-1500, 956-1425, 959-1662, 960-1430, 960-1586, 966-1678, 972-1487, 972-1500, 973-1187, 976-1576, 980-1470, 983-1520, 988-1750, 989-1562, 997-1525, 1002-1240, 1002-1275, 1005-1284, 1012-1269, 1016-1574, 1023-1465, 1024-1294, 1032-1305, 1033-1378, 1038-1653, 1045-1891, 1047-1258, 1052-1712, 1060-1773, 1063-1698, 1066-1286, 1076-1662, 1077-1537, 1081-1355, 1083-1344, 1086-1582, 1104-1708, 1105-1764, 1105-1790, 1111-1846, 1112-1697, 1132-1719, 1134-1381, 1137-1577, 1138-1404, 1140-1500, 1145-1716, 1155-1430, 1158-1691, 1163-1322, 1165-1468, 1166-1657, 1169-1464, 1177-1653, 1178-1413, 1179-1794, 1181-1619, 1191-1497, 1192-1444, 1192-1712, 1201-1691, 1205-1779, 1208-1465, 1213-1565, 1213-1572, 1213-1739, 1215-1690, 1239-1500,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
83 - Cont'd	1240-1552, 1241-1419, 1243-1535, 1244-1485, 1245-1377, 1251-1523, 1261-1641, 1265-1759, 1274-1531, 1276-1420, 1282-1551, 1284-1519, 1284-1524, 1297-1550, 1297-1581, 1311-1842, 1314-1584, 1314-1585, 1319-1561, 1322-1525, 1325-1623, 1325-1636, 1325-1783, 1326-1790, 1328-1641, 1332-1510, 1336-1858, 1337-1885, 1339-1469, 1339-1471, 1355-1610, 1355-1626, 1355-1801, 1363-1664, 1369-1622, 1369-1630, 1372-1681, 1372-1698, 1374-2233, 1375-1966, 1376-1630, 1384-1657, 1386-1678, 1388-1664, 1397-1612, 1399-1515, 1409-1512, 1421-1700, 1421-1989, 1446-1638, 1460-1777, 1494-1727, 1495-1682, 1498-1750, 1498-2117, 1506-1765, 1514-1882, 1526-1988, 1531-1799, 1535-2129, 1540-2227, 1545-1842, 1549-1644, 1549-1811, 1555-1820, 1563-1792, 1574-1872, 1580-2211, 1592-1862, 1593-2134, 1596-1906, 1599-1890, 1601-2181, 1607-2314, 1622-2211, 1634-2276, 1636-1907, 1636-1925, 1637-1925, 1638-2133, 1655-2024, 1667-2117, 1667-2130, 1667-2155, 1676-2215, 1678-1985, 1684-2106, 1690-2080, 1690-2313, 1706-2161, 1707-1942, 1707-2169, 1707-2326, 1713-2017, 1713-2259, 1716-2071, 1717-2026, 1718-2321, 1721-2251, 1723-2324, 1724-1876, 1728-2254, 1728-2268, 1730-1861, 1730-1940, 1730-1994, 1738-2277, 1740-2313, 1742-2299, 1743-2306, 1751-1789, 1754-2023, 1756-1919, 1757-2216, 1758-2000, 1758-2012, 1759-2272, 1760-2312, 1763-2048, 1763-2313, 1764-2320, 1767-2324, 1768-2260, 1769-2001, 1772-2021, 1776-2269, 1776-2328, 1783-2268, 1789-1985, 1789-2173, 1795-2057, 1795-2058, 1796-2047, 1796-2324, 1797-2038, 1805-2039, 1815-2212, 1815-2327, 1817-2073, 1819-2236, 1825-2077, 1825-2078, 1825-2236, 1826-2326, 1827-2327, 1828-2267, 1830-2267, 1834-2277, 1835-2331, 1838-2262, 1838-2291, 1848-1968, 1848-2071, 1850-2087, 1850-2119, 1851-1987, 1851-2261, 1851-2278; 1852-2308, 1853-2308, 1853-2309, 1854-2112, 1854-2235, 1855-2075, 1855-2146, 1855-2148, 1855-2310, 1856-2269; 1856-2343, 1858-2281, 1862-2106, 1862-2288, 1863-2334, 1865-2146, 1866-2219, 1866-2334, 1867-2282, 1871-2262, 1872-2321, 1874-2333, 1875-2297, 1878-2273, 1879-2148, 1879-2321, 1882-2309, 1887-2162, 1887-2170, 1887-2281, 1887-2282, 1890-2326, 1891-2307, 1892-2307, 1893-2313, 1894-2307, 1895-2128, 1895-2310, 1895-2354, 1896-2307, 1896-2308, 1896-2313, 1900-2049, 1900-2081, 1900-2204, 1900-2212, 1900-2308, 1900-2325, 1900-2344, 1901-2208, 1901-2295, 1902-2139, 1902-2305, 1902-2311, 1905-2162, 1910-2325, 1914-2305, 1915-2167, 1917-2185, 1917-2323, 1917-2327, 1920-2331, 1933-2306, 1934-2268, 1934-2271, 1936-2316, 1938-2307, 1939-2308, 1940-2310, 1941-2307, 1943-2307, 1944-2344, 1946-2307, 1947-2275, 1947-2311, 1949-2310, 1961-2330, 1963-2307, 1969-2307, 1971-2312, 1973-2308, 1974-2275, 1983-2321, 1988-2307, 1996-2344, 1997-2222, 1997-2288, 1998-2307, 2003-2307, 2004-2275, 2011-2309, 2011-2310, 2012-2309, 2017-2309, 2018-2289, 2019-2307, 2020-2307, 2028-2244, 2030-2255, 2036-2342, 2041-2307, 2042-2329, 2044-2307, 2046-2311, 2055-2333, 2056-2277, 2059-2307, 2083-2311, 2087-2307, 2092-2305, 2095-2312, 2095-2344, 2101-2321, 2101-2341, 2106-2310,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
83 - Cont'd	2109-2307, 2116-2315, 2118-2344, 2121-2344, 2129-2307, 2139-2307, 2140-2307, 2141-2321, 2150-2325, 2151-2321, 2152-2307, 2178-2319, 2194-2294, 2226-2336, 2228-2307, 2228-234
84/7581076CB1 2249	1-249, 3-499, 3-508, 3-2227, 166-741, 527-1073, 551-1150, 617-1138, 623-915, 678-846, 678-928, 683-1328, 784-1580, 830-1360, 884-1109, 890-1141, 915-1594, 920-1164, 943-1347, 961-1408, 969-1618, 981-1839, 987-1398, 992-1856, 1000-1407, 1006-1409, 1040-1319, 1044-1685, 1081-2090, 1083-1409, 1087-1282, 1144-1313, 1152-1840, 1168-2184, 1172-1313, 1260-1530, 1260-1690, 1282-1609, 1300-1561, 1300-1655, 1316-2243, 1344-1613, 1377-1792, 1458-1593, 1475-2092, 1526-2146, 1581-1850, 1581-2053, 1642-2064, 1653-2084, 1658-2068, 1665-1929, 1697-2249, 1716-2032, 1716-2061, 1738-2019, 1772-2063, 1778-2059
85/7504551CB1 1483	1-670, 11-264, 13-268, 14-1462, 18-254, 18-258, 18-293, 19-115, 19-613, 19-618, 19-640, 25-535, 38-115, 80-501, 110-288, 110-495, 110-590, 110-613, 126-405, 126-421, 128-369, 146-418, 148-849, 151-827, 153-823, 163-406, 166-1177, 168-454, 170-917, 174-252, 174-420, 181-726, 208-603, 210-387, 210-453, 216-502, 224-504, 239-330, 247-381, 255-505, 256-570, 261-640, 281-557, 292-522, 292-679, 296-534, 299-602, 302-702, 303-857, 308-847, 312-587, 327-519, 328-925, 332-638, 334-573, 339-640, 345-585, 349-670, 358-635, 360-1038, 365-663, 369-665, 376-817, 376-965, 390-989, 391-961, 397-1016, 400-682, 402-908, 402-1025, 404-659, 405-1036, 417-1092, 422-688, 425-1009, 426-728, 426-872, 430-1082, 432-589, 433-574, 433-702, 434-556, 434-739, 444-650, 454-789, 455-926, 458-696, 461-915, 463-717, 465-743, 479-913, 488-734, 491-1013, 494-756, 494-1030, 501-939, 501-1021, 502-802, 514-1025, 527-953, 528-774, 528-797, 532-816, 535-781, 541-919, 541-921, 545-1162, 553-800, 566-800, 567-768, 571-883, 572-1068, 579-1175, 589-831, 595-878, 607-902, 608-1196, 609-811, 609-1102, 611-884, 616-879, 632-911, 633-1053, 635-876, 646-915, 646-1251, 647-1306, 648-872, 653-952, 654-991, 656-682, 665-933, 666-844, 666-1046, 680-937, 683-1188, 687-945, 688-952, 688-963, 690-850, 691-1071, 692-1003, 703-934, 703-982, 711-814, 713-818, 713-958, 719-997, 719-1024, 723-1158, 723-1225, 737-999, 746-940, 746-1085, 748-957, 752-1080, 755-1029, 755-1065, 762-1378, 766-1011, 769-1299, 779-1019, 781-983, 781-1023, 781-1059, 784-1040, 784-1074, 784-1080, 785-1000, 785-1008, 788-1008, 789-1041, 797-1078, 800-1022, 820-1092, 822-1110, 824-1015, 844-1079, 844-1092, 848-1065, 851-1483, 853-1027, 853-1071, 853-1091, 853-1119, 853-1158, 853-1184, 853-1230, 853-1336, 853-1350, 853-1372, 853-1383, 853-1387, 853-1483, 855-1090, 860-1115, 860-1126, 869-1157, 874-1166, 876-1152, 877-1154, 877-1397, 878-1104, 879-1164, 882-1130, 885-1457, 886-1381, 889-1398, 893-1154, 894-1480, 896-1477, 898-1030, 904-1129, 904-1135, 904-1266, 906-1349,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
85 - Cont'd	909-1210, 910-1145, 910-1308, 913-1248, 917-1194, 919-1151, 922-1372, 926-1192, 931-1218, 934-1148, 938-1202, 940-1194, 942-1374, 952-1211, 965-1439, 966-1261, 976-1166, 976-1183, 978-1483, 979-1202, 991-1136, 991-1292, 994-1338, 1008-1250, 1008-1251, 1009-1254, 1016-1245, 1017-1447, 1021-1289, 1025-1190, 1028-1374, 1031-1325, 1035-1307, 1040-1374, 1042-1307, 1043-1315, 1043-1325, 1044-1374, 1051-1216, 1062-1374, 1065-1294, 1069-1338, 1070-1266, 1072-1348, 1072-1439, 1072-1482, 1072-1483, 1073-1372, 1074-1326, 1075-1374, 1088-1198, 1089-1390, 1092-1301, 1092-1387, 1092-1407, 1092-1411, 1109-1347, 1109-1349, 1113-1482, 1114-1483, 1118-1394, 1126-1367, 1126-1381, 1139-1403, 1148-1407, 1154-1380, 1159-1372, 1185-1411, 1185-1483, 1186-1439, 1238-1483, 1251-1462, 1263-1459, 1313-1374, 1413-1439
86/7500652CB1 1992	1-1992, 65-407, 349-550, 370-1014, 409-676, 439-536, 439-553, 439-600, 439-1042, 441-697, 441-1112, 447-636, 451-700, 454-727, 455-1110, 456-653, 456-656, 469-707, 486-1141, 507-1141, 516-770, 516-1026, 516-1036, 527-800, 527-1168, 541-719, 545-776, 548-637, 552-1112, 567-789, 574-985, 630-900, 646-1236, 647-1228, 652-807, 665-994, 670-804, 685-1104, 689-1300, 692-1232, 707-1263, 712-953, 727-995, 738-1069, 738-1071, 743-1043, 754-963, 757-1023, 759-1081, 760-1045, 764-992, 803-1335, 848-1453, 856-1222, 858-1134, 858-1317, 864-1315, 885-1159, 888-1136, 888-1149, 892-1133, 892-1301, 895-1266, 896-1565, 899-1332, 921-1160, 921-1453, 921-1532, 930-1222, 930-1391, 930-1606, 934-1459, 934-1512, 945-1151, 946-1481, 950-1139, 950-1208, 951-1218, 951-1230, 952-1159, 955-1221, 957-1617, 976-1221, 984-1244, 992-1566, 995-1273, 1014-1281, 1017-1255, 1038-1269, 1043-1205, 1044-1269, 1049-1314, 1050-1201, 1050-1284, 1051-1365, 1051-1474, 1051-1591, 1052-1254, 1052-1671, 1055-1303, 1055-1639, 1057-1308, 1057-1590, 1057-1836, 1062-1312, 1069-1317, 1072-1348, 1076-1269, 1079-1263, 1080-1328, 1080-1337, 1088-1561, 1093-1458, 1094-1261, 1094-1338, 1094-1579, 1101-1396, 1101-1399, 1102-1409, 1128-1839, 1135-1397, 1136-1426, 1145-1782, 1146-1269, 1151-1432, 1151-1436, 1157-1834, 1168-1379, 1175-1824, 1180-1417, 1180-1651, 1180-1840, 1183-1472, 1190-1833, 1200-1460, 1200-1523, 1204-1328, 1258-1548, 1260-1511, 1260-1665, 1264-1830, 1267-1530, 1267-1536, 1267-1598, 1272-1701, 1285-1541, 1295-1556, 1296-1560, 1297-1839, 1300-1839, 1301-1501, 1304-1586, 1304-1776, 1309-1830, 1315-1567, 1317-1839, 1319-1565, 1322-1527, 1327-1839, 1328-1559, 1328-1566, 1330-1839, 1334-1576, 1344-1584, 1350-1828, 1361-1601, 1361-1838, 1367-1840, 1382-1456, 1383-1518, 1385-1698, 1389-1838, 1395-1836, 1396-1839, 1401-1839, 1404-1610, 1404-1613, 1404-1667, 1407-1642, 1416-1711, 1416-1839, 1419-1840, 1427-1698, 1427-1828, 1427-1839, 1433-1841, 1436-1838, 1442-1631, 1442-1683, 1444-1836, 1447-1797, 1449-1838, 1449-1840, 1452-1839, 1456-1722, 1463-1837, 1464-1940, 1480-1790, 1482-1767, 1484-1839, 1494-1838, 1496-1838, 1497-1840, 1505-1839, 1506-1839, 1521-1794, 1564-1840, 1611-1839, 1661-1838, 1666-1912, 1683-1838, 1699-1969, 1725-1838, 1725-1992, 1728-1836, 1749-1979

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
87/7500900CB1 2597	1-524, 7-682, 7-731, 7-738, 13-257, 13-2597, 16-503, 16-776, 19-271, 21-243, 23-276, 23-317, 23-339, 32-847, 34-300, 34-721, 35-349, 35-762, 37-291, 40-812, 42-316, 49-303, 64-703, 64-704, 64-713, 64-751, 64-763, 64-813, 64-827, 64-829, 65-762, 89-372, 89-540, 89-749, 98-620, 107-729, 107-767, 126-745, 161-818, 164-830, 168-439, 183-494, 194-462, 215-463, 252-464, 256-838, 273-458, 273-517, 287-978, 295-902, 295-968, 305-864, 311-979, 313-830, 340-940, 361-596, 362-678, 374-575, 376-936, 474-701, 497-807, 497-1004, 501-978, 510-807, 525-1226, 526-836, 526-1219, 526-1253, 527-997, 556-789, 561-808, 586-833, 587-843, 595-1234, 599-1062, 617-1186, 671-1244, 675-871, 675-930, 694-1363, 696-971, 698-1363, 701-1210, 703-1342, 705-1363, 707-1329, 736-1278, 760-1363, 768-997, 783-1186, 804-1412, 821-1052, 822-1413, 884-1244, 888-1204, 907-1153, 917-1159, 919-1169, 919-1218, 920-1181, 920-1202, 933-1394, 985-1236, 989-1283, 991-1295, 995-1294, 1017-1292, 1035-1279, 1035-1290, 1035-1299, 1042-1321, 1042-1332, 1078-1521, 1078-1580, 1078-1599, 1096-1337, 1097-1387, 1098-1354, 1098-1376, 1126-1413, 1172-1401, 1173-1384, 1173-1413, 1187-1597, 1204-1395, 1211-1381, 1224-1362, 1345-2185, 1407-1596, 1407-1601, 1412-1697, 1412-1981, 1413-2178, 1467-1995, 1468-2185, 1472-1633, 1489-1737, 1490-1834, 1490-2011, 1494-2035, 1498-1732, 1508-2182, 1509-2253, 1509-2258, 1520-2183, 1533-2114, 1538-2177, 1559-2201, 1563-1819, 1569-2183, 1609-1830, 1631-1981, 1637-1893, 1668-2268, 1712-2328, 1716-2142, 1723-1981, 1723-2285, 1723-2325, 1739-1906, 1767-2068, 1767-2437, 1784-2443, 1791-2040, 1791-2471, 1795-2575, 1796-2307, 1796-2338, 1796-2345, 1796-2359, 1796-2382, 1796-2410, 1796-2415, 1796-2420, 1796-2430, 1796-2431, 1796-2440, 1796-2453, 1796-2455, 1796-2457, 1796-2495, 1796-2508, 1796-2512, 1796-2515, 1796-2538, 1796-2554, 1806-2086, 1809-2562, 1811-2066, 1812-2031, 1820-2037, 1830-2126, 1848-2493, 1857-2554, 1863-2453, 1871-2454, 1876-2214, 1882-2500, 1896-2179, 1896-2460, 1897-2521, 1901-2528, 1926-2173, 1926-2194, 1929-2195, 1930-2194, 1930-2398, 1945-2214, 1946-2469, 1946-2498, 1953-2576, 1967-2529, 1982-2259, 2016-2214, 2026-2538, 2029-2514, 2032-2165, 2045-2597, 2050-2180, 2050-2213, 2057-2185, 2057-2204, 2057-2291, 2057-2344, 2057-2586, 2058-2320, 2067-2355, 2068-2307, 2068-2330, 2096-2194, 2097-2356, 2101-2300, 2122-2402, 2133-2271, 2134-2287, 2158-2319, 2158-2413, 2226-2260
88/7501398CB1 1338	1-332, 20-1325, 181-796, 363-973, 417-941, 421-629, 424-679, 424-973, 483-737, 487-812, 505-747, 520-644, 535-755, 549-724, 557-818, 583-765, 610-779, 636-810, 664-916, 666-917, 688-814, 694-1323, 706-967, 709-963, 712-973, 727-906, 727-975, 727-1286, 732-855, 740-1282, 752-1275, 871-1282, 875-1274, 969-1243, 969-1248, 969-1272, 969-1282, 969-1286, 971-1338, 972-1243, 987-1191, 987-1279, 995-1192, 1002-1275, 1013-1275, 1019-1264, 1024-1213, 1033-1286, 1034-1247, 1036-1275, 1038-1282, 1045-1282, 1052-1282, 1069-1282, 1084-1282, 1122-1286, 1126-1282, 1138-1264, 1163-1282, 1196-1272

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
89/7501417CB1 1705	1-240, 1-1638, 1-1705, 26-786, 46-272, 69-705, 69-724, 403-599, 407-599, 408-981, 411-572, 416-552, 447-1034, 448-1034, 462-1407, 470-1401, 522-1401, 523-1401, 575-1253, 589-1350, 591-1401, 604-1401, 659-1274, 666-899, 702-1401, 750-1401, 849-1401, 902-1401, 949-1598, 985-1622, 1062-1600, 1087-1368, 1161-1654, 1182-1628, 1205-1629, 1218-1510, 1218-1628, 1218-1705, 1221-1347, 1254-1598, 1254-1630, 1282-1628, 1356-1630, 1404-1630
90/7501472CB1 1623	1-1591, 173-393, 174-444, 191-454, 229-502, 232-469, 253-485, 283-513, 283-617, 570-902, 857-1172, 857-1212, 857-1262, 857-1327, 857-1358, 941-1095, 965-1191, 1036-1532, 1133-1591, 1154-1358, 1162-1505, 1334-1591, 1358-1591, 1375-1591, 1441-1533, 1441-1623
91/7501489CB1 673	1-242, 5-278, 12-259, 13-194, 13-296, 13-316, 14-257, 14-314, 16-251, 16-333, 17-314, 18-263, 18-265, 18-276, 18-294, 18-670, 20-288, 20-365, 21-261, 22-250, 22-261, 23-287, 24-225, 24-227, 24-300, 24-302, 24-306, 25-265, 25-318, 29-279, 29-289, 29-321, 31-261, 31-309, 32-237, 32-273, 32-296, 32-335, 37-262, 37-477, 39-177, 47-310, 48-287, 48-301, 48-311, 48-321, 49-258, 51-275, 51-302, 51-311, 53-299, 53-332, 53-345, 55-354, 56-302, 56-307, 56-313, 59-311, 59-347, 74-211, 78-185, 116-243, 129-268, 171-421, 176-441, 202-452, 217-673, 237-662, 316-673, 377-673, 379-656, 389-673, 410-611, 410-638, 428-673, 429-668
92/7501555CB1 1088	1-248, 20-127, 20-540, 21-157, 21-306, 23-101, 26-279, 28-332, 31-203, 31-246, 31-271, 31-465, 31-487, 31-526, 31-557, 31-642, 33-618, 34-314, 36-267, 37-357, 41-317, 42-311, 44-312, 50-127, 52-327, 70-378, 116-356, 116-381, 116-416, 128-339, 128-344, 128-398, 128-419, 128-736, 130-426, 131-1069, 135-668, 139-357, 141-372, 141-668, 142-369, 157-240, 157-360, 166-354, 167-440, 171-748, 173-456, 193-439, 193-455, 193-460, 194-321, 194-547, 208-741, 219-746, 220-701, 233-719, 244-530, 257-732, 260-527, 260-728, 274-449, 277-541, 277-552, 285-520, 307-526, 309-611, 338-612, 340-601, 348-629, 364-578, 386-617, 393-663, 423-683, 447-694, 447-1067, 448-669, 479-712, 479-728, 479-745, 479-939, 522-783, 591-1088, 652-1084, 666-1081, 676-1077, 698-1088, 736-977, 749-1077, 750-993, 750-1061, 751-1088, 756-1077, 759-1004, 759-1037, 759-1077, 775-1079, 792-1077, 822-1085, 837-1077, 839-1075, 839-1077, 846-1068, 848-1077, 854-1077, 857-1086, 858-1088, 859-1044, 862-1077, 897-1088, 908-1025, 916-1088

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
93/7501561CB1 1401	1-139, 1-169, 1-237, 1-278, 1-364, 1-426, 1-448, 1-464, 1-472, 1-473, 1-495, 1-507, 1-522, 1-559, 1-587, 1-617, 1-619, 1-623, 1-630, 1-643, 1-646, 1-658, 1-684, 1-709, 1-758, 1-1395, 2-227, 20-249, 23-584, 25-250, 32-242, 34-138, 50-303, 132-830, 220-962, 232-764, 254-651, 254-1043, 274-500, 274-617, 274-784, 274-848, 274-876, 274-931, 291-644, 294-543, 296-1035, 309-1013, 340-991, 357-910, 358-544, 358-624, 361-890, 379-519, 380-948, 381-519, 385-1021, 397-686, 411-1069, 417-1281, 422-656, 436-1164, 448-998, 451-713, 456-1142, 462-1178, 464-725, 498-1084, 504-1379, 509-1027, 520-1147, 536-1159, 558-1166, 560-706, 560-828, 582-1194, 583-817, 602-1260, 605-941, 617-1252, 622-1101, 623-879, 627-1354, 628-1180, 630-884, 631-1194, 632-878, 634-1130, 639-849, 639-1275, 640-1143, 657-1274, 658-1158, 660-1284, 663-1330, 665-1206, 665-1318, 666-1206, 667-1241, 683-1234, 686-1300, 687-988, 693-1305, 699-1338, 713-1369, 730-1313, 739-1226, 739-1279, 739-1348, 749-974, 750-1189, 750-1315, 756-1080, 765-1047, 767-1215, 782-924, 782-1271, 795-1313, 795-1320, 805-1356, 814-1013, 816-959, 821-953, 821-1106, 827-1171, 830-1360, 832-1398, 835-1057, 840-1323, 846-1306, 854-1083, 855-1336, 861-1147, 862-1289, 870-1163, 872-1071, 880-978, 881-980, 903-1259, 904-1162, 907-1400, 911-1330, 917-1244, 928-1162, 946-1082, 987-1279, 994-1215, 1001-1084, 1012-1208, 1061-1355, 1068-1317, 1071-1326, 1074-1334, 1079-1353, 1081-1283, 1089-1320, 1089-1344, 1102-1359, 1127-1401, 1143-1401, 1169-1376, 1169-1401
94/7506108CB1 1656	1-274, 1-1656, 5-287, 15-300, 15-344, 25-301, 71-322, 82-245, 85-387, 108-317, 158-373, 299-866, 308-363, 311-394, 328-925, 393-583, 393-605, 393-609, 393-620, 393-739, 393-750, 393-833, 393-924, 393-952, 393-960, 393-1146, 398-582, 404-696, 413-675, 413-1069, 414-655, 418-807, 421-674, 421-1069, 431-679, 432-658, 433-692, 435-692, 435-1069, 439-1061, 445-1029, 451-968, 467-959, 473-703, 484-628, 487-605, 487-754, 487-769, 490-811, 492-1158, 494-947, 496-758, 496-908, 497-1156, 499-1024, 505-770, 506-694, 507-755, 507-785, 507-794, 518-767, 519-1076, 523-803, 526-739, 526-811, 529-775, 529-1211, 530-1054, 530-1195, 531-1131, 534-715, 535-1173, 537-819, 542-853, 546-1141, 547-840, 547-1389, 548-950, 552-1155, 560-820, 561-783, 561-1158, 564-804, 568-916, 569-1063, 569-1213, 571-838, 571-1149, 574-864, 574-1205, 575-675, 576-840, 576-871, 576-1358, 577-1112, 579-1205, 580-916, 582-961, 585-1107, 587-874, 588-1140, 590-826, 597-1077, 600-836, 608-858, 608-965, 609-992, 622-990, 623-1084, 623-1184,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
94 - Cont'd	625-878, 633-861, 640-1149, 641-889, 650-1282, 652-916, 652-1207, 654-919, 660-837, 660-912, 662-1007, 663-922, 663-926, 663-1206, 663-1266, 664-916, 664-1332, 665-908, 668-1323, 672-916, 672-1436, 674-1272, 678-964, 679-1285, 681-1394, 685-1335, 686-908, 687-1115, 687-1242, 689-1343, 690-1214, 692-1110, 695-935, 696-970, 700-1080, 701-916, 701-933, 701-940, 701-965, 701-1094, 703-1193, 706-984, 706-986, 709-1400, 711-980, 711-1045, 711-1287, 712-1156, 713-970, 718-972, 721-958, 728-964, 728-975, 737-1233, 739-1040, 740-1066, 743-1012, 743-1053, 747-1206, 748-1136, 758-1225, 761-1007, 764-1218, 764-1358, 769-987, 769-1330, 770-1065, 774-916, 776-1084, 779-964, 783-1008, 783-1196, 786-1202, 793-1026, 795-1048, 796-982, 798-1330, 805-1038, 805-1066, 810-1231, 813-1220, 824-1499, 828-1335, 832-1275, 833-1656, 835-1304, 843-979, 844-969, 845-1114, 850-1057, 852-1368, 854-1121, 866-1038, 866-1363, 866-1364, 869-1129, 869-1168, 874-1114, 877-1147, 879-1038, 888-1194, 895-1163, 897-1233, 903-1162, 904-1169, 909-1391, 918-1038, 919-1358, 922-1465, 940-1248, 942-1234, 946-1185, 946-1230, 946-1259, 946-1518, 952-1261, 955-1455, 976-1304, 977-1260, 982-1201, 982-1246, 987-1221, 988-1221, 988-1229, 1004-1287, 1008-1243, 1008-1301, 1013-1224, 1013-1281, 1020-1267, 1020-1288, 1023-1374, 1027-1217, 1027-1305, 1032-1153, 1033-1299, 1035-1272, 1035-1279, 1035-1283, 1039-1204, 1039-1291, 1039-1348, 1040-1142, 1040-1510, 1041-1385, 1041-1432, 1042-1292, 1050-1239, 1055-1331, 1059-1307, 1059-1314, 1059-1321, 1062-1332, 1066-1330, 1067-1291, 1067-1525, 1080-1221, 1097-1502, 1099-1351, 1099-1364, 1099-1371, 1100-1516, 1102-1379, 1114-1516, 1117-1463, 1120-1320, 1120-1390, 1124-1340, 1143-1410, 1145-1384, 1147-1303, 1147-1364, 1148-1423, 1153-1386, 1159-1325, 1161-1409, 1164-1428, 1168-1367, 1171-1428, 1179-1450, 1183-1393, 1183-1443, 1184-1426, 1185-1371, 1185-1432, 1186-1447, 1196-1456, 1207-1467, 1209-1359, 1214-1466, 1236-1495, 1265-1513, 1267-1488, 1287-1421
95/7506123CB1 1841	1-206, 1-221, 1-237, 1-243, 1-245, 1-247, 1-254, 1-258, 1-287, 1-591, 1-652, 1-716, 1-834, 1-835, 1-947, 2-258, 3-258, 3-654, 3-1841, 9-247, 9-258, 10-163, 10-258, 11-229, 11-245, 12-258, 15-258, 20-258, 22-224, 22-234, 30-258, 35-258, 40-258, 50-934, 75-231, 91-258, 96-936, 183-1028, 259-501, 259-690, 263-907, 280-1273, 292-1054, 294-571, 310-784, 313-756, 314-976, 315-488, 321-548, 339-975, 357-610, 365-626, 365-641, 367-651, 371-1269, 375-732, 380-640, 385-674, 385-886, 390-694, 390-737, 390-859, 390-871, 396-671, 403-652, 403-672, 409-882, 426-762, 428-881, 440-695, 446-668, 446-1147, 502-695, 511-751, 514-636, 530-785, 538-786, 539-792, 542-992, 547-992, 572-813, 577-851, 583-1186, 596-889, 598-826, 601-869, 601-888, 601-993, 601-1072, 619-913, 621-1098, 625-955, 626-1055, 633-949, 634-1336, 637-869, 637-1143, 638-888, 654-1045, 673-890, 673-1210, 688-984, 699-1290, 700-1184, 702-988, 719-1020, 722-972, 725-1204, 725-1617, 734-1246, 738-921, 738-957, 738-980, 738-1000, 738-1278, 744-1366,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
95 - Cont'd	745-969, 745-1289, 750-1104, 754-1438, 755-1152, 758-1439, 769-1215, 774-972, 775-1616, 777-1171, 783-944, 783-1351, 787-1035, 789-1112, 791-1269, 805-1586, 812-1066, 837-1472, 841-1617, 855-1518, 861-1061, 861-1495, 866-1404, 876-1362, 877-1590, 883-1268, 888-1617, 890-1471, 896-1140, 900-1128, 902-1416, 903-1176, 903-1566, 907-1566, 909-1184, 910-1179, 916-1565, 927-1537, 927-1580, 927-1616, 928-1595, 939-1224, 944-1237, 947-1617, 956-1235, 956-1617, 960-1187, 973-1115, 974-1215, 974-1522, 974-1617, 976-1617, 977-1214, 977-1223, 978-1186, 978-1532, 978-1562, 979-1617, 981-1570, 984-1566, 1000-1184, 1015-1614, 1038-1617, 1042-1279, 1047-1373, 1051-1562, 1054-1569, 1055-1617, 1057-1366, 1080-1617, 1084-1355, 1084-1369, 1085-1617, 1091-1327, 1091-1617, 1094-1349, 1094-1566, 1096-1487, 1099-1456, 1104-1365, 1109-1362, 1110-1395, 1122-1250, 1133-1351, 1139-1617, 1142-1387, 1154-1617, 1157-1398, 1164-1425, 1165-1420, 1166-1414, 1169-1617, 1170-1558, 1173-1617, 1175-1617, 1179-1617, 1182-1617, 1184-1617, 1186-1617, 1189-1418, 1189-1617, 1190-1617, 1192-1447, 1193-1446, 1194-1468, 1194-1617, 1195-1497, 1199-1617, 1200-1607, 1204-1617, 1212-1617, 1213-1617, 1222-1617, 1223-1493, 1223-1617, 1227-1617, 1231-1617, 1233-1617, 1238-1467, 1248-1504, 1257-1617, 1267-1497, 1271-1555, 1276-1617, 1278-1617, 1279-1617, 1280-1617, 1283-1497, 1284-1617, 1286-1433, 1286-1617, 1291-1617, 1298-1383, 1298-1497, 1298-1524, 1312-1617, 1315-1617, 1321-1617, 1325-1617, 1333-1617, 1335-1553, 1337-1616, 1345-1603, 1347-1617, 1348-1490, 1351-1617, 1358-1830, 1364-1617, 1369-1498, 1372-1617, 1376-1607, 1384-1617, 1409-1617, 1413-1617, 1422-1617, 1444-1617, 1452-1617, 1464-1617, 1466-1617, 1475-1617, 1476-1617, 1515-1617, 1609-1673, 1609-1712, 1609-1713, 1609-1722, 1609-1724, 1609-1736, 1609-1744, 1609-1766, 1609-1775, 1609-1779, 1609-1804, 1609-1812, 1609-1816, 1609-1824, 1609-1830, 1617-1830, 1618-1830, 1619-1830, 1620-1830, 1621-1830, 1622-1830, 1623-1830, 1626-1829, 1626-1830, 1630-1830, 1633-1830, 1635-1830, 1651-1830, 1656-1830, 1664-1830, 1666-1830, 1670-1830, 1684-1830, 1690-1819, 1691-1830, 1693-1830, 1695-1830, 1698-1830, 1701-1830, 1716-1830, 1717-1830, 1720-1830, 1721-1830, 1732-1830, 1741-1830, 1742-1830, 1749-1830, 1750-1830, 1755-1830, 1763-1830, 1768-1830, 1770-1830, 1772-1830, 1774-1830, 1784-1827, 1790-1830, 1793-1830, 1801-1830, 1805-1830

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
96/7506248CB1 3659	1-793, 1-843, 5-271, 14-308, 14-509, 14-608, 14-3659, 28-571, 28-843, 29-624, 29-862, 30-526, 38-797, 55-337, 55-703, 55-864, 55-919, 61-618, 63-900, 71-845, 73-362, 84-874, 91-755, 101-543, 118-362, 178-467, 253-518, 326-578, 367-976, 435-1209, 441-695, 472-989, 477-1270, 488-1154, 503-794, 512-1099, 555-990, 585-1050, 597-1038, 597-1171, 601-1168, 602-1121, 651-1336, 655-981, 676-1273, 691-1325, 735-1520, 752-1007, 765-983, 786-1453, 867-1153, 935-1468, 968-1571, 1028-1177, 1036-1930, 1068-1571, 1100-1522, 1102-1384, 1144-1373, 1146-1571, 1217-1741, 1236-1571, 1236-1830, 1291-1571, 1313-1529, 1346-1571, 1415-1571, 1568-1618, 1568-1712, 1568-1888, 1568-1963, 1572-1808, 1595-1863, 1606-2340, 1623-1874, 1660-2197, 1673-2404, 1690-1952, 1697-2221, 1729-1978, 1759-2493, 1773-2000, 1774-1971, 1792-2281, 1811-2598, 1844-2417, 1877-2153, 1882-2403, 1884-2358, 1886-2577, 1894-2193, 1901-2463, 1904-2639, 1906-2534, 1929-2751, 1932-2592, 1939-2715, 1957-2584, 1960-2454, 1964-2486, 1964-2513, 2021-2706, 2026-2360, 2026-2711, 2028-2355, 2033-2362, 2053-2338, 2077-2594, 2090-2748, 2105-2348, 2113-2368, 2135-2643, 2136-2258, 2136-2373, 2145-2412, 2150-2436, 2153-2628, 2166-2662, 2184-2369, 2189-2707, 2189-2731, 2199-2728, 2208-2457, 2209-3107, 2213-2449, 2227-2456, 2233-3107, 2234-2481, 2248-2469, 2255-3107, 2255-3111, 2260-3107, 2262-2708, 2263-2731, 2263-2738, 2264-2934, 2305-2983, 2342-2888, 2347-2600, 2359-2628, 2359-3107, 2360-2798, 2362-3107, 2371-2631, 2374-2929, 2383-3017, 2385-2983, 2390-2960, 2393-2649, 2399-2703, 2410-2665, 2426-2713, 2443-2677, 2451-2892, 2451-2964, 2451-2985, 2451-3207, 2451-3210, 2454-3030, 2470-2739, 2471-2989, 2481-2704, 2481-2705, 2481-2719, 2488-2727, 2500-2766, 2512-3185, 2519-3127, 2520-2744, 2552-2864, 2554-2809, 2572-3222, 2577-3049, 2581-2848, 2584-2812, 2596-3150, 2614-2924, 2615-2778, 2636-2838, 2637-2976, 2646-3246, 2653-3107, 2653-3111, 2660-2948, 2664-2916, 2666-3160, 2679-3139, 2682-2929, 2704-2958, 2707-3316, 2717-3257, 2721-2989, 2728-2968, 2768-3295, 2790-3042, 2793-3071, 2795-3471, 2797-3051, 2798-2980, 2798-3027, 2804-3304, 2805-3043, 2813-3261, 2814-3521, 2854-3159, 2859-3120, 2863-3076, 2875-3102, 2881-3239, 2882-3499, 2896-3160, 2928-3494, 2928-3582, 2935-3177, 2944-3571, 2954-3117, 2964-3565, 2966-3273, 2967-3244, 2985-3300, 2986-3294, 2993-3583, 2998-3257, 3006-3257, 3007-3521, 3008-3499, 3009-3613, 3010-3130, 3019-3588, 3021-3579, 3023-3616, 3024-3600, 3033-3519, 3035-3643, 3036-3302, 3036-3552, 3046-3293, 3049-3655, 3050-3521, 3054-3550, 3058-3216, 3058-3323, 3077-3323, 3092-3323, 3093-3323, 3094-3323, 3100-3323, 3109-3323, 3117-3337, 3127-3557, 3127-3610, 3145-3323, 3145-3417, 3146-3404, 3147-3334, 3148-3569, 3149-3410, 3150-3496, 3155-3655, 3157-3456, 3159-3653, 3173-3568, 3182-3659, 3186-3352, 3190-3416, 3197-3414, 3197-3444, 3199-3431, 3203-3479, 3203-3584, 3205-3445, 3205-3454, 3205-3457, 3205-3461, 3205-3485, 3205-3487, 3205-3535, 3208-3646, 3225-3415, 3225-3454, 3229-3644, 3229-3657, 3231-3655, 3232-3437,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
96 - Cont'd	3244-3507, 3245-3504, 3246-3504, 3253-3445, 3254-3535, 3258-3659, 3265-3657, 3268-3655, 3270-3659, 3273-3659, 3276-3659, 3313-3554, 3318-3591, 3323-3381, 3323-3659, 3324-3485, 3324-3539, 3324-3552, 3324-3567, 3324-3570, 3324-3579, 3324-3580, 3324-3581, 3324-3609, 3345-3572, 3346-3608, 3351-3609, 3352-3506, 3363-3581, 3364-3521, 3366-3617, 3379-3644, 3381-3590, 3386-3654, 3401-3654, 3407-3645, 3407-3652, 3407-3659, 3422-3655, 3424-3655, 3432-3650, 3459-3588, 3493-3647
97/7506347CB1 1935	1-228, 1-386, 1-470, 1-507, 1-1935, 44-609, 44-659, 96-546, 129-528, 130-615, 145-546, 166-699, 170-744, 171-600, 171-734, 184-834, 242-843, 259-618, 328-872, 345-951, 348-796, 348-807, 379-835, 379-958, 382-1010, 394-961, 397-1007, 402-1080, 404-983, 414-916, 439-744, 439-834, 439-895, 454-834, 458-930, 486-782, 509-1138, 530-1062, 531-1020, 547-978, 547-1078, 556-1078, 574-1173, 579-1174, 609-1239, 620-790, 621-1128, 622-1100, 655-954, 665-1197, 669-1254, 687-1305, 703-1141, 703-1154, 707-966, 715-1319, 729-1300, 745-1214, 752-1371, 770-1135, 812-1455, 825-1021, 825-1391, 839-1365, 841-1474, 843-1086, 844-1475, 856-1118, 856-1470, 873-1499, 876-1273, 879-1430, 928-1075, 961-1434, 967-1182, 967-1540, 969-1194, 984-1509, 990-1434, 993-1523, 1011-1481, 1014-1663, 1016-1291, 1023-1294, 1027-1660, 1041-1390, 1044-1444, 1045-1671, 1070-1705, 1074-1481, 1076-1649, 1087-1388, 1088-1315, 1143-1663, 1145-1612, 1158-1424, 1184-1758, 1191-1430, 1191-1649, 1198-1388, 1215-1668, 1227-1730, 1234-1535, 1245-1740, 1266-1886, 1268-1535, 1277-1595, 1280-1592, 1284-1539, 1292-1822, 1294-1766, 1301-1935, 1315-1935, 1354-1935, 1376-1878, 1377-1846, 1377-1873, 1410-1935, 1437-1921, 1457-1935, 1463-1759, 1463-1935, 1469-1794, 1469-1935, 1471-1935, 1494-1935, 1495-1935, 1496-1935, 1499-1935, 1504-1935, 1505-1935, 1507-1915, 1512-1921, 1514-1935, 1515-1935, 1516-1935, 1518-1931, 1518-1935, 1519-1935, 1521-1935, 1523-1935, 1524-1935, 1526-1935, 1538-1935, 1539-1935, 1540-1935, 1542-1935, 1547-1935, 1552-1794, 1557-1935, 1558-1746, 1584-1935, 1598-1833, 1625-1935, 1626-1935, 1628-1935, 1630-1935, 1651-1935, 1662-1935, 1667-1935, 1673-1935, 1679-1935, 1683-1935, 1721-1935, 1757-1935
98/7509172CB1 2240	1-252, 1-790, 1-2131, 3-351, 4-684, 6-334, 6-339, 6-354, 13-199, 13-226, 14-236, 509-1087, 602-817, 625-1266, 689-1208, 721-1269, 723-945, 723-969, 723-1035, 723-1147, 754-1162, 754-1163, 770-1270, 877-1782, 921-1169, 921-1175, 1119-1486, 1127-1521, 1127-1522, 1217-1803, 1441-2009, 1468-1908, 1474-2147, 1484-2042, 1495-2147, 1503-2031, 1517-1891, 1542-2083, 1567-2070, 1570-1805, 1570-2105, 1570-2109, 1574-2182, 1592-2097, 1592-2173, 1607-1866, 1607-2133, 1610-1829, 1630-2059, 1630-2135, 1635-2195, 1661-2240, 1679-2237, 1685-1888, 1686-1888, 1698-2159, 1768-2133, 1814-1965, 1881-2135

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
99/7510421CB1 1812	1-255, 1-259, 1-1806, 152-439, 152-787, 152-962, 167-995, 329-842, 390-1069, 413-971, 419-617, 426-670, 431-683, 437-692, 439-975, 503-725, 628-1225, 676-915, 732-1106, 761-1261, 843-1194, 876-1520, 877-1732, 911-1605, 915-1441, 915-1444, 933-1689, 939-1417, 942-1180, 946-1426, 955-1210, 957-1189, 996-1787, 1008-1547, 1009-1538, 1011-1544, 1013-1601, 1032-1212, 1038-1627, 1042-1291, 1044-1329, 1045-1225, 1045-1293, 1052-1267, 1057-1579, 1066-1576, 1069-1352, 1072-1521, 1075-1599, 1075-1649, 1086-1262, 1086-1366, 1086-1385, 1086-1640, 1111-1375, 1113-1383, 1117-1373, 1119-1614, 1127-1779, 1134-1562, 1144-1414, 1146-1429, 1162-1628, 1163-1406, 1166-1447, 1166-1732, 1172-1403, 1177-1757, 1187-1537, 1187-1651, 1205-1758, 1215-1488, 1216-1505, 1220-1743, 1231-1784, 1239-1806, 1240-1806, 1245-1474, 1263-1812, 1265-1507, 1265-1812, 1266-1462, 1266-1690, 1266-1727, 1273-1559, 1286-1528, 1293-1543, 1293-1576, 1302-1514, 1306-1720, 1307-1458, 1312-1590, 1320-1812, 1322-1812, 1323-1796, 1324-1757, 1329-1797, 1338-1812, 1341-1804, 1342-1643, 1343-1812, 1344-1810, 1346-1795, 1347-1786, 1347-1793, 1348-1794, 1350-1756, 1352-1607, 1353-1645, 1358-1812, 1359-1789, 1362-1812, 1364-1640, 1364-1646, 1366-1806, 1368-1792, 1371-1642, 1371-1788, 1373-1798, 1375-1792, 1376-1793, 1377-1556, 1377-1788, 1380-1793, 1380-1799, 1380-1808, 1381-1789, 1381-1795, 1382-1780, 1382-1795, 1383-1788, 1385-1799, 1385-1812, 1386-1658, 1386-1753, 1386-1792, 1386-1797, 1392-1631, 1392-1793, 1392-1798, 1393-1792, 1397-1597, 1397-1790, 1397-1793, 1398-1640, 1398-1650, 1398-1793, 1399-1792, 1406-1793, 1407-1792, 1408-1803, 1409-1788, 1409-1796, 1412-1794, 1415-1688, 1415-1693, 1419-1778, 1420-1658, 1420-1792, 1422-1792, 1423-1724, 1427-1796, 1443-1675, 1443-1711, 1443-1712, 1443-1792, 1443-1793, 1445-1795, 1447-1778, 1451-1636, 1452-1805, 1452-1806, 1459-1630, 1459-1705, 1459-1719, 1459-1753, 1459-1792, 1474-1791, 1474-1808, 1479-1525, 1481-1675, 1484-1793, 1486-1793, 1486-1795, 1492-1696, 1492-1812, 1496-1794, 1497-1763, 1497-1775, 1498-1788, 1499-1795, 1504-1792, 1507-1795, 1519-1792, 1524-1776, 1524-1802, 1526-1803, 1543-1796, 1544-1802, 1550-1812, 1553-1789, 1553-1806, 1557-1804, 1562-1775, 1562-1790, 1566-1792, 1570-1792, 1571-1796, 1575-1796, 1577-1792, 1579-1784, 1586-1812, 1587-1792, 1588-1792, 1590-1792, 1590-1795, 1593-1812, 1596-1812, 1600-1784, 1600-1791, 1603-1718, 1620-1789, 1625-1812, 1639-1795, 1650-1812, 1677-1806, 1687-1795, 1690-1811, 1693-1788, 1698-1804

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
100/7504625CB1 995	1-247, 1-392, 14-248, 14-254, 14-259, 14-266, 14-385, 17-294, 20-990, 21-287, 33-278, 33-285, 41-385, 42-301, 46-302, 50-348, 51-304, 54-344, 55-310, 55-339, 55-363, 55-392, 56-309, 57-339, 72-350, 75-357, 75-392, 84-345, 84-861, 85-361, 90-311, 94-313, 102-383, 103-374, 106-287, 125-391, 148-380, 203-456, 203-687, 212-968, 224-820, 224-944, 249-984, 361-618, 398-995, 460-986, 479-990, 479-995, 511-749, 514-995, 521-990, 528-990, 533-990, 537-812, 540-989, 547-990, 550-995, 562-990, 568-989, 569-990, 576-752, 576-990, 589-988, 590-995, 604-988, 615-989, 615-993, 627-862, 627-874, 630-988, 645-985, 646-990, 656-990, 658-986, 659-990, 667-995, 669-990, 675-990, 676-990, 686-989, 718-990, 728-986, 756-990, 759-990, 764-990, 781-990, 793-989, 799-988, 819-990, 857-995, 910-990
101/7504776CB1 1297	1-515, 1-658, 5-290, 11-270, 13-283, 13-316, 13-322, 14-255, 15-255, 16-190, 16-200, 16-247, 16-252, 16-256, 16-260, 16-272, 16-287, 16-296, 16-309, 16-311, 16-313, 16-320, 16-325, 16-1297, 18-262, 18-274, 19-236, 19-300, 19-308, 21-274, 21-298, 21-300, 21-309, 21-310, 21-317, 21-325, 23-353, 25-330, 26-296, 26-305, 30-297, 30-331, 32-155, 32-214, 32-253, 32-259, 32-264, 32-276, 32-280, 32-289, 32-316, 32-319, 32-320, 32-321, 32-325, 32-328, 32-330, 32-333, 32-343, 32-353, 33-293, 35-190, 35-298, 36-341, 37-343, 40-239, 42-345, 46-347, 48-263, 70-353, 112-351, 175-463, 214-495, 214-686, 352-558, 352-563, 352-566, 352-593, 352-601, 352-604, 352-610, 352-625, 352-630, 352-665, 352-689, 352-836, 352-914, 352-926, 352-1003, 353-559, 354-623, 358-633, 361-549, 364-627, 365-594, 366-589, 369-669, 370-680, 374-959, 378-669, 378-690, 383-555, 385-613, 385-628, 387-1033, 391-677, 393-689, 393-789, 395-1018, 398-633, 399-941, 404-637, 405-640, 405-651, 405-793, 408-714, 410-564, 410-708, 413-702, 415-604, 415-623, 415-635, 415-686, 415-1049, 421-891, 421-1090, 425-697, 429-1101, 431-1031, 434-571, 434-661, 434-668, 435-670, 436-582, 436-660, 436-786, 436-891, 437-1073, 440-1056, 440-1064, 443-740, 446-613, 453-1228, 453-1289, 453-1297, 454-692, 455-615, 455-823, 455-941, 458-764, 461-731, 462-712, 465-970, 468-1128, 469-712, 470-700, 470-740, 472-761, 474-1244, 478-772, 480-791, 482-718, 485-905, 488-1240, 489-734, 490-746, 490-1208, 491-999, 494-789, 498-789, 500-740, 504-779, 506-729, 506-745, 506-1144, 510-738, 510-1113, 511-763, 511-778, 511-806, 513-749, 526-772, 529-1205, 536-701, 536-753, 536-1087, 542-854, 542-1291, 551-786, 556-1194, 556-1216, 560-1268, 563-1135, 570-1176, 574-824, 574-827, 574-830, 574-1293, 575-1293, 582-829, 586-826, 587-838, 587-889, 588-834, 589-869, 593-914, 593-929, 593-967, 593-1097, 593-1175, 593-1180, 593-1195, 593-1289, 598-741, 598-824, 598-862, 598-870, 601-1167, 605-956, 605-1194, 607-849, 612-1081, 615-864, 615-866, 616-850, 616-853, 616-973,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
101 - Cont'd	624-865, 626-1056, 626-1200, 626-1287, 628-1141, 640-881, 642-861, 643-863, 643-934, 645-891, 645-936, 649-1262, 650-817, 650-884, 653-1278, 663-923, 663-1204, 667-937, 674-1194, 675-1279, 680-912, 680-1297, 689-919, 690-938, 695-921, 695-933, 702-1293, 706-1207, 707-934, 710-1202, 713-1217, 716-1297, 718-1123, 718-1220, 718-1280, 719-1244, 720-1248, 727-1035, 728-960, 728-979, 731-976, 735-954, 742-974, 744-1011, 744-1028, 744-1060, 745-1284, 750-1297, 755-1055, 755-1255, 757-969, 757-981, 757-1028, 760-1006, 760-1030, 760-1077, 762-1032, 762-1062, 762-1095, 762-1279, 762-1296, 763-1001, 764-984, 767-938, 769-1011, 771-953, 771-982, 771-1003, 771-1014, 778-905, 779-1062, 786-1278, 791-1040, 792-1034, 792-1038, 792-1102, 798-1051, 801-937, 805-1282, 806-1061, 807-1083, 814-1153, 816-991, 818-1256, 818-1296, 819-1297, 825-1294, 835-1070, 835-1290, 850-1101, 852-1294, 854-1291, 857-1292, 859-1084, 861-1294, 862-1038, 862-1292, 866-1049, 866-1294, 867-1086, 872-1162, 876-965, 881-1105, 881-1297, 883-1294, 885-1200, 889-1139, 889-1193, 890-1290, 891-1117, 891-1140, 893-1103, 893-1162, 894-1292, 898-1202, 900-1296, 905-1157, 905-1162, 905-1293, 905-1294, 907-1109, 909-1148, 914-1297, 919-1236, 930-1158, 932-1167, 933-1130, 938-1191, 939-1292, 945-1202, 945-1209, 955-1254, 966-1292, 967-1214, 969-1248, 970-1266, 972-1211, 972-1217, 972-1246, 972-1297, 982-1232, 984-1293, 986-1213, 986-1226, 986-1287, 986-1288, 986-1289, 986-1291, 986-1292, 986-1293, 986-1294, 986-1296, 986-1297, 989-1248, 990-1198, 990-1210, 991-1214, 991-1218, 992-1240, 1000-1207, 1000-1297, 1004-1294, 1010-1236, 1011-1224, 1021-1188, 1024-1294, 1036-1294, 1037-1297, 1039-1216, 1055-1259, 1059-1289, 1059-1290, 1059-1297, 1060-1294, 1062-1292, 1067-1294, 1070-1293, 1071-1290, 1073-1280, 1075-1290, 1077-1297, 1082-1297, 1084-1294, 1084-1297, 1089-1294, 1091-1297, 1103-1287, 1103-1294, 1104-1280, 1114-1297, 1117-1281, 1136-1294, 1138-1294, 1138-1297, 1148-1294, 1166-1280, 1166-1294, 1171-1294, 1192-1297, 1194-1266, 1205-1297, 1206-1297, 1215-1297
102/7504927CB1 600	1-276, 1-600, 24-311, 44-129, 44-295, 48-152, 65-299, 88-299, 252-460, 252-592, 256-600, 452-589, 453-600, 509-599

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
103/7505010CB1 2292	1-251, 1-258, 1-273, 1-304, 1-377, 1-402, 1-460, 1-491, 1-506, 1-608, 2-1701, 14-534, 37-576, 39-290, 52-537, 80-637, 95-586, 136-584, 145-295, 145-351, 145-381, 145-451, 145-467, 145-598, 145-644, 145-656, 145-710, 145-750, 145-792, 145-911, 145-923, 146-704, 150-391, 151-403, 152-380, 152-769, 153-389, 153-415, 156-496, 156-780, 157-862, 157-954, 162-391, 163-434, 163-435, 163-470, 163-590, 163-951, 164-351, 164-428, 164-979, 165-696, 166-443, 166-479, 167-570, 167-739, 168-518, 168-546, 169-434, 170-464, 171-413, 171-437, 172-804, 172-1093, 173-583, 173-604, 173-755, 173-756, 178-405, 178-411, 178-449, 178-475, 178-642, 178-748, 178-758, 178-797, 178-865, 178-866, 178-918, 179-729, 182-949, 182-1113, 185-530, 192-537, 192-780, 193-441, 193-949, 196-475, 199-458, 199-780, 199-798, 199-859, 200-467, 200-485, 200-661, 201-481, 201-528, 204-792, 205-401, 205-516, 205-1025, 206-491, 206-599, 210-752, 215-428, 217-560, 218-950, 225-703, 230-873, 247-519, 247-522, 252-899, 254-1026, 259-748, 262-538, 287-848, 320-812, 320-1018, 327-502, 371-1071, 377-657, 377-707, 383-1008, 383-1078, 388-671, 389-673, 390-930, 406-526, 415-835, 415-1149, 426-1028, 431-576, 431-1149, 434-1116, 441-1032, 449-1055, 450-936, 454-966, 463-712, 472-1099, 475-735, 477-737, 487-722, 513-784, 513-982, 515-846, 539-827, 539-866, 539-895, 541-804, 543-1009, 545-794, 545-809, 552-738, 556-934, 556-1111, 589-859, 601-1122, 621-815, 640-1039, 643-856, 643-878, 680-930, 680-932, 691-892, 708-940, 712-956, 716-938, 718-1438, 725-833, 743-1016, 771-1008, 782-908, 785-951, 788-1056, 789-1025, 790-1050, 793-1130, 828-1092, 838-1052, 838-1082, 845-950, 856-1117, 863-1063, 879-1121, 887-1134, 934-1094, 953-1168, 1028-1461, 1038-1270, 1140-1438, 1148-1441, 1151-1819, 1165-1317, 1175-1447, 1192-1415, 1211-1761, 1223-1440, 1234-1441, 1238-1581, 1298-1557, 1330-1939, 1373-1758, 1374-1675, 1415-1687, 1443-1688, 1449-1627, 1461-1696, 1476-1631, 1484-1777, 1499-1670, 1505-1701, 1506-1790, 1523-1687, 1542-2292, 1584-1667, 1609-1761
104/7505173CB1 789	1-124, 1-191, 1-553, 5-162, 5-186, 5-271, 5-530, 7-136, 7-159, 7-311, 8-148, 18-106, 18-191, 21-162, 21-191, 23-551, 27-133, 198-529, 211-365, 294-523, 319-789

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
105/7510061CB1 3702	1-599, 1-3659, 65-719, 163-896, 163-951, 163-969, 163-1028, 163-1105, 163-1111, 165-1118, 166-1136, 169-937, 235-558, 245-679, 264-649, 285-710, 571-865, 584-1182, 657-840, 658-902, 658-1030, 658-1092, 658-1220, 658-1257, 658-1335, 658-1351, 691-932, 691-1254, 756-1252, 766-1330, 790-1305, 800-1362, 809-1554, 832-1177, 856-1119, 869-1440, 871-1520, 881-1152, 881-1346, 917-1429, 1010-1673, 1015-1263, 1018-1046, 1018-1048, 1046-1242, 1059-1242, 1182-1921, 1201-1908, 1309-1430, 1345-2052, 1347-1615, 1372-2060, 1388-1970, 1399-1681, 1403-1870, 1404-2105, 1434-2176, 1446-2114, 1459-1547, 1459-2163, 1479-2129, 1479-2148, 1479-2172, 1479-2185, 1480-2148, 1480-2181, 1486-2357, 1487-1735, 1502-1813, 1510-2116, 1556-1900, 1557-2144, 1563-1868, 1574-2106, 1587-1873, 1659-1908, 1659-2185, 1659-2245, 1691-1908, 1734-1980, 1734-2213, 1744-2008, 1766-2271, 1802-2031, 1836-2126, 1847-2125, 1892-2381, 1930-2239, 1930-2354, 1967-2218, 1983-2213, 1983-2235, 1983-2456, 1983-2515, 1983-2611, 1983-2660, 1986-2677, 2037-2599, 2073-2330, 2074-2281, 2255-2490, 2255-2510, 2255-2845, 2262-2558, 2272-2594, 2324-2827, 2435-3087, 2457-2845, 2465-3050, 2552-2878, 2590-3219, 2626-2930, 2646-3029, 2647-3258, 2672-2927, 2673-2914, 2674-2939, 2676-3287, 2699-2995, 2699-3138, 2731-2999, 2732-3337, 2733-3006, 2734-3215, 2747-3357, 2747-3358, 2783-3040, 2783-3362, 2796-3183, 2806-3008, 2806-3281, 2810-3153, 2811-3059, 2811-3272, 2815-3441, 2833-3410, 2836-3418, 2844-3339, 2844-3400, 2845-3364, 2850-3258, 2860-3636, 2868-3638, 2870-3074, 2875-3286, 2885-3483, 2892-3092, 2892-3121, 2892-3185, 2892-3234, 2892-3374, 2892-3463, 2894-3129, 2899-3520, 2903-3162, 2903-3478, 2905-3187, 2905-3188, 2905-3555, 2907-3164, 2911-3470, 2912-3398, 2927-3539, 2927-3657, 2931-3406, 2931-3484, 2946-3433, 2952-3283, 2963-3203, 2963-3568, 2969-3534, 2971-3169, 2975-3640, 2976-3231, 2977-3383, 2978-3617, 2982-3599, 2986-3502, 2986-3614, 2986-3630, 2995-3595, 2997-3264, 2997-3528, 3000-3508, 3000-3587, 3001-3483, 3016-3642, 3022-3626, 3024-3332, 3033-3607, 3040-3367, 3041-3478, 3043-3499, 3043-3598, 3045-3459, 3061-3408, 3079-3655, 3090-3601, 3095-3587, 3096-3640, 3103-3635, 3105-3659, 3108-3286, 3114-3640, 3120-3429, 3128-3665, 3135-3659, 3135-3664, 3137-3571, 3139-3413, 3139-3417, 3147-3639, 3150-3589, 3154-3654, 3161-3640, 3170-3640, 3173-3640, 3174-3640, 3180-3639, 3183-3639, 3185-3430, 3185-3464, 3185-3631, 3185-3640, 3191-3659, 3191-3661, 3192-3640, 3195-3640, 3207-3664, 3210-3654, 3211-3640, 3218-3640, 3220-3628, 3220-3640, 3220-3660, 3221-3640, 3231-3640, 3235-3640, 3236-3671, 3236-3702, 3239-3534, 3239-3640, 3240-3643, 3240-3644, 3244-3640, 3245-3640, 3249-3655, 3254-3639, 3262-3640, 3274-3640, 3278-3660, 3284-3441, 3285-3640, 3293-3640, 3295-3644, 3300-3642, 3314-3491, 3323-3640, 3332-3578, 3334-3538, 3346-3660, 3362-3640, 3367-3640, 3380-3483, 3380-3606, 3380-3628, 3402-3618, 3408-3640, 3422-3640, 3427-3640, 3461-3640, 3461-3702, 3465-3640, 3466-3640, 3470-3599, 3476-3640, 3527-3640, 3527-3655

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
106/7510091CB1 2060	1-173, 1-204, 1-272, 1-273, 1-275, 1-281, 1-282, 1-2060, 2-278, 3-261, 28-256, 30-244, 37-281, 49-278, 54-256, 55-246, 55-281, 60-256, 66-281, 67-256, 76-256, 272-532, 283-943, 283-958, 317-673, 317-688, 317-696, 317-754, 318-795, 328-821, 328-822, 328-826, 328-827, 328-834, 328-857, 328-862, 376-636, 378-621, 378-794, 513-805, 539-771, 765-991, 796-997, 1913-2060, 1929-2060
107/7510109CB1 3692	1-284, 1-710, 1-3692, 27-417, 29-550, 29-690, 39-303, 39-539, 45-673, 189-648, 207-473, 408-877, 415-891, 615-934, 754-1225, 1784-2047, 2035-2515, 2494-2588, 2494-2679, 2494-2857, 2494-2968, 2494-2983, 2494-2998, 2494-3039, 2494-3040, 2494-3048, 2494-3084, 2494-3136, 2494-3155, 2494-3166, 2494-3175, 2494-3180, 2494-3190, 2494-3210, 2494-3219, 2496-2729, 2499-3189, 2614-2882, 2614-2896, 2667-2907, 2668-2863, 2688-2909, 2688-3106, 2688-3117, 2704-3398, 2825-3072, 2825-3271, 2850-3123, 2912-3294, 2918-3535, 2926-3411, 2936-3217, 2936-3436, 2947-3655, 3117-3383, 3209-3672, 3350-3692
108/7510121CB1 2088	1-217, 1-269, 1-270, 1-296, 1-595, 2-227, 2-297, 2-2083, 3-221, 3-286, 4-291, 5-289, 5-292, 5-610, 6-340, 7-222, 7-487, 8-245, 8-297, 9-273, 9-620, 10-220, 10-261, 10-271, 10-426, 10-453, 10-554, 13-219, 13-257, 13-261, 13-584, 20-265, 20-567, 20-588, 22-688, 27-524, 28-700, 30-543, 34-170, 34-301, 34-508, 34-614, 34-676, 34-683, 35-275, 35-281, 35-307, 35-322, 36-295, 37-345, 37-679, 38-274, 38-314, 38-316, 38-356, 38-549, 38-599, 39-226, 39-283, 41-288, 41-320, 41-643, 41-673, 41-676, 41-706, 42-152, 42-306, 44-292, 44-297, 44-331, 45-289, 46-338, 47-312, 47-319, 47-326, 56-596, 59-303, 59-642, 60-241, 62-322, 63-334, 71-308, 73-347, 120-372, 125-229, 125-394, 144-489, 199-600, 219-496, 221-417, 225-488, 232-493, 234-399, 234-555, 238-675, 245-461, 249-555, 265-555, 279-477, 293-566, 298-649, 305-546, 306-544, 327-561, 329-577, 362-534, 373-666, 378-655, 394-606, 400-702, 433-516, 441-705, 451-663, 684-953, 686-1292, 853-1402, 857-1301, 870-1380, 911-1541, 927-1185, 929-1199, 930-1266, 937-1217, 937-1613, 944-1586, 974-1171, 974-1234, 976-1235, 978-1227, 986-1265, 1004-1683, 1006-1383, 1006-1596, 1007-1590, 1015-1135, 1015-1250, 1018-1595, 1022-1300, 1029-1198, 1054-1643, 1074-1202, 1100-1491, 1100-1586, 1100-1647, 1109-1612, 1119-1762, 1120-1622, 1127-1603, 1131-1727, 1153-1348, 1153-1762, 1172-1866, 1175-1816, 1177-1777, 1177-1824, 1192-1833, 1198-1827, 1201-1824, 1222-1815, 1224-1603, 1224-1775, 1246-1869, 1272-1869, 1274-1854, 1276-1735, 1281-2021, 1290-2030, 1297-1890, 1300-1890, 1301-2022, 1302-1872, 1311-1890, 1313-1738, 1313-1790, 1313-1836, 1314-2049, 1317-1847, 1328-2020, 1328-2031, 1331-1805, 1349-1532, 1355-1990, 1356-1941, 1369-1970, 1377-1971, 1377-1972, 1377-2049, 1379-1795, 1379-1817, 1379-1831, 1379-1866, 1384-1890, 1385-2030, 1391-2021, 1392-1890, 1392-1969, 1396-1883, 1408-1597, 1413-1753, 1416-1908, 1425-1883, 1431-1886, 1431-2088, 1435-1984, 1435-2022, 1446-2001, 1453-1969, 1457-1886, 1457-1890, 1466-1883, 1467-2049, 1472-1948, 1477-1730, 1480-1989,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
108 - Cont'd	1481-1691, 1481-2072, 1487-1662, 1488-1779, 1488-1787, 1488-1869, 1497-1944, 1504-1904, 1507-1709, 1515-2079, 1517-1746, 1517-1749, 1522-2052, 1527-1883, 1529-1852, 1529-2088, 1548-1791, 1553-1827, 1559-2083, 1577-1795, 1578-1838, 1578-1856, 1617-1880, 1624-1882, 1647-1880, 1653-2080, 1667-1937, 1673-1973, 1679-1938, 1679-2072, 1680-1869, 1686-2072, 1686-2077, 1686-2081, 1695-2072, 1699-2072, 1701-1916, 1702-2079, 1705-1987, 1706-1995, 1711-1977, 1714-2081, 1723-1988, 1732-2072, 1739-2068, 1745-2080, 1753-2029, 1754-2086, 1757-1953, 1764-2020, 1798-2072, 1804-2082, 1821-2075, 1831-2072, 1853-2072, 1882-2088, 1884-1953, 1985-2072, 2045-2078
109/7510797CB1 3020	1-279, 1-3020, 62-198, 106-492, 142-377, 147-397, 321-609, 517-879, 522-876, 552-695, 560-1230, 567-1198, 614-1398, 642-1249, 667-1067, 696-1254, 699-861, 722-1148, 724-1176, 731-1164, 740-1287, 748-1203, 754-1358, 755-1345, 766-1255, 781-1470, 792-1474, 826-1451, 833-1243, 833-1250, 841-1217, 846-1364, 847-1351, 862-1271, 868-1338, 875-1489, 881-1549, 884-1210, 893-1139, 912-1486, 912-1501, 913-1523, 915-1200, 915-1205, 967-1619, 1013-1434, 1070-1632, 1072-1527, 1073-1727, 1082-1726, 1085-1407, 1110-1462, 1117-1687, 1118-1414, 1135-1886, 1138-1760, 1139-1579, 1139-1611, 1142-1804, 1152-1513, 1163-1894, 1206-1682, 1214-1861, 1216-1791, 1222-1874, 1224-1599, 1241-1674, 1261-1717, 1261-2047, 1262-1662, 1266-1506, 1266-1537, 1266-1708, 1266-1777, 1266-1791, 1266-1802, 1266-1815, 1266-1817, 1266-1845, 1266-1874, 1266-1935, 1266-1936, 1266-1977, 1266-1990, 1266-2018, 1266-2022, 1268-1851, 1269-1980, 1270-1918, 1277-1935, 1279-1887, 1290-1902, 1297-1691, 1299-1963, 1301-1827, 1301-1882, 1328-1878, 1329-2041, 1334-1487, 1355-1988, 1369-1952, 1379-1722, 1379-1724, 1379-1863, 1379-1892, 1379-1990, 1379-2051, 1385-2134, 1394-1828, 1395-1920, 1402-1898, 1431-1633, 1432-1726, 1432-2051, 1456-2015, 1480-1735, 1508-2173, 1512-1777, 1512-1960, 1513-2309, 1514-2034, 1540-2080, 1545-2207, 1562-2116, 1567-1762, 1611-2252, 1618-2088, 1627-2243, 1631-1855, 1637-2241, 1639-1866, 1658-2289, 1676-2323, 1680-2287, 1691-1871, 1692-2010, 1693-1964, 1695-1889, 1695-2049, 1698-2292, 1715-2253, 1720-2037, 1725-1971, 1727-2034, 1727-2258, 1779-2375, 1780-2568, 1783-2415, 1808-2642, 1817-2077, 1840-1923, 1841-2346, 1846-2372, 1853-2495, 1873-2437, 1874-2166, 1895-2538, 1896-2538, 1921-2618, 1923-2637, 1926-2106, 1931-2222, 1931-2226, 1934-2187, 1934-2624, 1944-2624, 1950-2565, 1954-2278, 1958-2228, 1962-2616, 1966-2657, 1971-2371, 1979-2306, 1982-2449, 1985-2656, 1999-2656, 2003-2436, 2004-2654, 2005-2550, 2009-2654, 2013-2656, 2027-2463, 2037-2371, 2040-2641, 2049-2295, 2050-2535, 2051-2656, 2054-2533,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
109 - Cont'd	2059-2288, 2059-2514, 2061-2244, 2069-2573, 2081-2656, 2091-2601, 2129-2652, 2135-2676, 2143-2364, 2143-2651, 2151-2656, 2160-2309, 2163-2656, 2168-2699, 2169-2610, 2174-2599, 2175-2656, 2176-2656, 2185-2656, 2187-2656, 2188-2656, 2194-2657, 2196-2641, 2217-2641, 2224-2705, 2232-2642, 2241-2656, 2250-2656, 2257-2640, 2258-2656, 2259-2595, 2268-2606, 2279-2656, 2297-2995, 2321-2656, 2322-2915, 2339-2499, 2342-2640, 2343-2641, 2344-2877, 2345-2600, 2349-2807, 2352-2907, 2352-2928, 2352-2966, 2352-3011, 2353-2656, 2369-2973, 2375-2587, 2384-2656, 2385-2656, 2392-2632, 2398-2643, 2402-2547, 2406-2656, 2410-2899, 2423-2685, 2423-2972, 2459-2645, 2517-2883, 2557-2883, 2563-3002, 2563-3006, 2577-2839, 2577-2999, 2579-3002, 2595-2869, 2599-2866, 2609-3005, 2619-3020, 2633-2879, 2657-2876, 2664-2922, 2667-3006, 2667-3007, 2667-3011, 2674-2930, 2682-3015, 2690-3007, 2694-3020, 2696-3005, 2696-3010, 2697-3020, 2724-2979, 2745-2933, 2791-2952, 2806-2886, 2835-3003, 2837-3018, 2848-3020
110/7504944CB1 2109	1-466, 10-224, 11-230, 14-222, 15-222, 18-764, 18-893, 18-2032, 19-635, 36-697, 41-191, 44-232, 47-190, 47-232, 47-629, 58-689, 75-338, 118-760, 176-338, 225-652, 225-663, 225-678, 225-707, 225-714, 227-659, 231-814, 233-688, 234-542, 234-550, 234-566, 234-698, 234-707, 234-778, 234-784, 234-907, 243-710, 245-869, 255-667, 257-467, 258-668, 267-904, 277-641, 293-563, 298-580, 319-664, 319-682, 323-550, 340-698, 366-944, 394-657, 408-859, 410-659, 410-855, 413-972, 417-629, 436-674, 440-819, 440-824, 445-1037, 447-1044, 484-761, 487-1044, 491-1052, 507-1201, 509-1108, 550-796, 552-1195, 555-1151, 556-1138, 558-824, 558-825, 558-826, 564-798, 565-1092, 565-1164, 569-1149, 584-1135, 615-1154, 633-921, 639-841, 641-1338, 642-1149, 653-1186, 658-1280, 663-929, 664-879, 665-860, 665-942, 667-807, 672-922, 673-917, 675-942, 675-992, 678-903, 682-915, 686-1291, 688-1197, 709-1306, 714-942, 720-980, 720-1103, 720-1156, 720-1187, 720-1219, 720-1564, 734-957, 757-1185, 772-1384, 775-1288, 777-1293, 785-1484, 798-1116, 812-1068, 817-1476, 825-1047, 827-992, 827-1084, 827-1398, 840-1051, 844-1531, 844-1546, 888-1454, 918-1460, 920-1476, 923-1479, 923-1486, 924-1455, 927-1493, 932-1277, 945-1308, 957-1207, 957-1214, 968-1221, 975-1334, 978-1217, 991-1629, 992-1056, 1003-1241, 1007-1254, 1008-1493, 1020-1288, 1025-1585, 1026-1292, 1044-1569, 1050-1715, 1054-1727, 1055-1303, 1058-1275, 1081-1295, 1083-1539, 1084-1281, 1100-1225, 1102-1203, 1102-1222, 1127-1367, 1134-1647, 1142-1422, 1150-2024, 1153-1425, 1157-1420, 1161-1718, 1163-1474, 1167-1383, 1175-1435, 1177-1578, 1178-1792, 1184-1384, 1197-1782, 1206-1583, 1212-1834, 1213-2035, 1219-1503, 1220-1463, 1222-1859, 1224-1822, 1225-1826, 1229-1493, 1229-1959, 1229-1965, 1229-1980, 1229-1985, 1229-1989, 1229-1993, 1229-1994, 1229-2033, 1229-2046, 1229-2086, 1232-1903, 1232-2032, 1246-1989, 1247-2027, 1252-1784, 1263-1790, 1263-1892, 1265-2033, 1269-1507, 1269-1514, 1269-1973, 1273-1545, 1284-1979, 1287-1491, 1287-1505, 1287-1518, 1292-1526,

Table 4

Polynucleotide SEQ ID NO:/ Incyte ID/ Sequence Length	Sequence Fragments
110 - Cont'd	1295-1957, 1304-1671, 1306-1548, 1312-1913, 1325-1608, 1336-1907, 1337-1905, 1340-1918, 1350-1552, 1350-1586, 1354-1956, 1356-1552, 1371-1614, 1372-1580, 1398-2004, 1398-2020, 1398-2021, 1401-1659, 1404-1655, 1404-1660, 1404-1690, 1417-1666, 1417-1712, 1432-1856, 1444-1729, 1455-1727, 1458-1982, 1466-2029, 1467-1693, 1467-1712, 1468-1785, 1483-2027, 1483-2089, 1493-2035, 1494-1652, 1494-1804, 1500-1969, 1500-2045, 1501-1742, 1508-1718, 1510-1700, 1510-2035, 1511-1989, 1514-1761, 1520-2035, 1524-1993, 1525-1757, 1527-1792, 1534-1801, 1559-1638, 1560-2027, 1561-2035, 1562-1859, 1570-1846, 1575-1821, 1577-2059, 1579-2038, 1586-1873, 1587-1772, 1588-2040, 1589-2035, 1591-1841, 1594-1838, 1594-2044, 1595-2038, 1597-1826, 1597-2033, 1600-1860, 1600-2024, 1601-2059, 1601-2089, 1602-1845, 1611-2044, 1612-1965, 1616-2078, 1616-2089, 1617-1904, 1618-2035, 1621-2057, 1630-2042, 1631-2042, 1636-1918, 1640-1838, 1640-1884, 1640-2042, 1642-2029, 1644-1949, 1647-1843, 1654-2030, 1654-2035, 1654-2036, 1654-2040, 1654-2042, 1659-1901, 1676-1928, 1678-1950, 1685-2036, 1687-2035, 1688-1926, 1690-2042, 1694-1888, 1709-1946, 1711-1985, 1713-1978, 1720-2035, 1721-2028, 1730-2035, 1741-1972, 1743-2017, 1744-1930, 1764-2035, 1769-1996, 1769-2042, 1774-2019, 1774-2032, 1778-2022, 1794-2056, 1799-2035, 1808-2057, 1814-2032, 1815-2029, 1824-2071, 1824-2086, 1838-2086, 1842-2109, 1860-2038, 1865-2090, 1902-2030, 1906-2086, 1921-2074, 1935-2026, 1935-2036, 1956-2029, 2072-2109

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID:	Representative Library
56	7509350CB1	HEAONOC01
57	7509325CB1	LIVRNOT01
58	7509337CB1	KIDNNOT19
59	7509353CB1	LIVRTUT04
60	7509354CB1	BSTMNON02
61	7509385CB1	BSTMNON02
63	7509376CB1	LATRTUT02
64	7501927CB1	SINTNOT19
65	7503274CB1	BRAHTDK01
66	7509104CB1	BRAVUNT02
67	7509996CB1	ADRENOT07
68	7510030CB1	THYMFET02
69	7510062CB1	PITUNON01
70	7510217CB1	SMCCNON03
71	7510298CB1	NERDTDN03
72	7510299CB1	SINTNOR01
73	7510368CB1	PROTDNV28
74	7510369CB1	BLADTUT07
75	7510377CB1	BRSTNOT14
76	7510026CB1	BRSTNOT12
77	7509168CB1	ADRENOT07
78	7500607CB1	LVENNOT03
79	7506079CB1	FIBRTXS07
80	7509259CB1	LIVRDIT02
81	7509263CB1	THP1NOT03
82	7509360CB1	ADRENOT07
83	7509394CB1	ADRETUE02
84	7581076CB1	BRADDIR01
85	7504551CB1	PROSTUT12
86	7500652CB1	SINTNOR01

Table 5

Polynucleotide SEQ ID NO:	Incyte Project ID:	Representative Library
87	7500900CB1	PHOSDNV44
88	7501398CB1	NERDTDN03
89	7501417CB1	LIVRFET02
90	7501472CB1	LIVRTUT13
91	7501489CB1	BRAINOT03
92	7501555CB1	KERANOT01
93	7501561CB1	KERANOT02
94	7506108CB1	PROSTMY01
95	7506123CB1	BRAYDIT01
96	7506248CB1	PITUDIR01
97	7506347CB1	SYNORAT03
98	7509172CB1	LIVRNOT01
99	7510421CB1	SPLNNOT02
100	7504625CB1	COLNFET02
101	7504776CB1	ADRENOT07
102	7504927CB1	COLNNOT23
103	7505010CB1	TESTTUT02
104	7505173CB1	ENDANOT01
105	7510061CB1	BRSTNOT17
106	7510091CB1	BMARTXE01
107	7510109CB1	BRAITUT26
108	7510121CB1	PROSBPT02
109	7510797CB1	LATRNOT01
110	7504944CB1	SINTFER02

Table 6

Library	Vector	Library Description
ADRENOT07	pINCY	Library was constructed using RNA isolated from adrenal tissue removed from a 61-year-old female during a bilateral adrenalectomy. Patient history included an unspecified disorder of the adrenal glands.
ADRETUE02	PCDNA2.1	This 5' biased random primed library was constructed using RNA isolated from right adrenal tumor tissue removed from a 49-year-old Caucasian male during unilateral adrenalectomy. Pathology indicated adrenal cortical carcinoma comprising nearly the entire specimen. The tumor was attached to the adrenal gland which showed mild cortical atrophy. The tumor was encapsulated, being surrounded by a thin (1-3 mm) rim of connective tissue. The patient presented with adrenal cancer, abdominal pain, pyrexia of unknown origin, and deficiency anemia. Patient history included benign hypertension. Previous surgeries included adenotonsillectomy. Patient medications included aspirin, calcium, and iron. Family history included atherosclerotic coronary artery disease in the mother; cerebrovascular accident and atherosclerotic coronary artery disease in the father; and benign hypertension in the grandparent(s).
BLADTUT07	pINCY	Library was constructed using RNA isolated from bladder tumor tissue removed from the anterior bladder wall of a 58-year-old Caucasian male during a radical cystectomy, radical prostatectomy, and gastrostomy. Pathology indicated a grade 3 transitional cell carcinoma in the left lateral bladder. Patient history included angina, emphysema, and tobacco use. Family history included acute myocardial infarction, atherosclerotic coronary artery disease, and type II diabetes.
BMARTXE01	pINCY	This 5' biased random primed library was constructed using RNA isolated from treated SH-SY5Y cells derived from a metastatic bone marrow neuroblastoma, removed from a 4-year-old Caucasian female (Schering AG). The medium was MEM/HAM'S F12 with 10% fetal calf serum. After reaching about 80% confluency cells were treated with 6-Hydroxydopamine (6-OHDA) at 100 microM for 8 hours.
BRADDIR01	pINCY	Library was constructed using RNA isolated from diseased choroid plexus tissue of the lateral ventricle, removed from the brain of a 57-year-old Caucasian male, who died from a cerebrovascular accident.

Table 6

Library	Vector	Library Description
BRAHTDK01	PSPORT1	This amplified and normalized library was constructed using pooled RNA isolated from archaecortex, anterior and posterior hippocampus tissue removed from a 55-year-old Caucasian female who died from cholangiocarcinoma. Pathology indicated mild meningeal fibrosis predominately over the convexities, scattered axonal spheroids in the white matter of the cingulate cortex and the thalamus, and a few scattered neurofibrillary tangles in the entorhinal cortex and the periaqueductal gray region. Pathology for the associated tumor tissue indicated well-differentiated cholangiocarcinoma of the liver with residual or relapsed tumor. Patient history included cholangiocarcinoma, post-operative Budd-Chiari syndrome, biliary ascites, hydrothorax, dehydration, malnutrition, oliguria and acute renal failure. Previous surgeries included cholecystectomy and resection of 85% of the liver. 7.6x10e5 independent clones from this amplified library were normalized in 1 round using conditions adapted Soares et al., PNAS (1994) 91:9228-9232 and Bonaldo et al., Genome Research (1996) 6:791, except that a significantly longer (48 hours/round) reannealing hybridization was used.
BRAINOT03	PSPORT1	Library was constructed using RNA isolated from brain tissue removed from a 26-year-old Caucasian male during cranioplasty and excision of a cerebral meningeal lesion. Pathology for the associated tumor tissue indicated a grade 4 oligoastrocytoma in the right fronto-parietal part of the brain.
BRAITUT26	pINCY	Library was constructed using RNA isolated from brain tumor tissue removed from the right posterior fossa, occipital convexity of a 70-year-old Caucasian male during cerebral meninges lesion excision. Pathology indicated meningioma. Patient history included a benign colon neoplasm and unspecified personality disorder. Family history included chronic proliferative nephritis, acute myocardial infarction, atherosclerotic coronary artery disease, and chronic proliferative nephritis.
BRAVUNT02	PSPORT1	Library was constructed using pooled RNA isolated from separate populations of unstimulated astrocytes.
BRAYDIT01	pINCY	Library was constructed using RNA isolated from diseased hypothalamus tissue removed from a 57-year-old Caucasian male who died from a cerebrovascular obstruction. Patient history included Huntington's disease and emphysema.
BRSTNOT12	pINCY	Library was constructed using RNA isolated from diseased breast tissue removed from a 32-year-old Caucasian female during a bilateral reduction mammoplasty. Pathology indicated nonproliferative fibrocystic disease. Family history included benign hypertension and atherosclerotic coronary artery disease.

Table 6

Library	Vector	Library Description
BRSTNOT14	pINCY	Library was constructed using RNA isolated from breast tissue removed from a 62-year-old Caucasian female during a unilateral extended simple mastectomy. Pathology for the associated tumor tissue indicated an invasive grade 3 (of 4), nuclear grade 3 (of 3) adenocarcinoma, ductal type. Ductal carcinoma in situ, comedo type, comprised 60% of the tumor mass. Metastatic adenocarcinoma was identified in one (of 14) axillary lymph nodes with no perinodal extension. The tumor cells were strongly positive for estrogen receptors and weakly positive for progesterone receptors. Patient history included a benign colon neoplasm, hyperlipidemia, cardiac dysrhythmia, and obesity. Family history included atherosclerotic coronary artery disease, myocardial infarction, colon cancer, ovarian cancer, lung cancer, and cerebrovascular disease.
BRSTNOT17	pINCY	Library was constructed using RNA isolated from breast tissue removed from a 46-year-old Caucasian female during a unilateral extended simple mastectomy. Pathology for the associated tumor tissue indicated invasive grade 3, nuclear grade 2 adenocarcinoma, ductal type. An intraductal carcinoma component, non-comedo, comprised approximately 50% of the neoplasm, including the lactiferous ducts. Angiolymphatic involvement was present, and metastatic adenocarcinoma was present in 7 of 10 axillary lymph nodes. The largest nodal metastasis measured 3 cm, and focal extracapsular extension was identified. Family history included atherosclerotic coronary artery disease, type II diabetes, cerebrovascular disease, and depressive disorder.
BSTMNON02	PSPORT1	This normalized brain stem library was constructed from 2.84 million independent clones from a brain stem library. Starting RNA was made from the brain stem tissue of a 72-year-old Caucasian male who died from myocardial infarction. Patient history included coronary artery disease, insulin-dependent diabetes mellitus, and arthritis. Normalization and hybridization conditions were adapted from Soares et al. (PNAS (1994) 91:9228).
COLNFET02	pINCY	Library was constructed using RNA isolated from the colon tissue of a Caucasian female fetus, who died at 20 weeks' gestation.
COLNNOT23	pINCY	Library was constructed using RNA isolated from diseased colon tissue removed from a 16-year-old Caucasian male during a total colectomy with abdominal/perineal resection. Pathology indicated gastritis and pancolonitis consistent with the acute phase of ulcerative colitis. Inflammation was more severe in the transverse colon, with inflammation confined to the mucosa. There was only mild involvement of the ascending and sigmoid colon, and no significant involvement of the cecum, rectum, or terminal ileum. Family history included irritable bowel syndrome.
ENDANOT01	PBLUESCRIPT	Library was constructed using RNA isolated from aortic endothelial cell tissue from an explanted heart removed from a male during a heart transplant.

Table 6

Library	Vector	Library Description
FIBRTXS07	pINCY	This subtracted library was constructed using 1.3 million clones from a dermal fibroblast library and was subjected to two rounds of subtraction hybridization with 2.8 million clones from an untreated dermal fibroblast tissue library. The starting library for subtraction was constructed using RNA isolated from treated dermal fibroblast tissue removed from the breast of a 31-year-old Caucasian female. The cells were treated with 9CIS retinoic acid. The hybridization probe for subtraction was derived from a similarly constructed library from RNA isolated from untreated dermal fibroblast tissue from the same donor. Subtractive hybridization conditions were based on the methodologies of Swaroop et al., NAR (1991) 19:1954 and Bonaldo, et al., Genome Research (1996) 6:791.
HEAONOC01	PSPORT1	This large size fractionated library was constructed using RNA isolated from the aorta of a 39-year-old Caucasian male, who died from a gunshot wound. Serology was positive for cytomegalovirus (CMV). Patient history included tobacco abuse (one pack of cigarettes per day for 25 years), and occasionally cocaine, marijuana, and alcohol use.
KERANOT01	PBLUESCRIPT	Library was constructed using RNA isolated from neonatal keratinocytes obtained from the leg skin of a spontaneously aborted black male.
KERANOT02	PSPORT1	Library was constructed using RNA isolated from epidermal breast keratinocytes (NHEK). NHEK (Clontech #CC-2501) is a human breast keratinocyte cell line derived from a 30-year-old black female during breast-reduction surgery.
KIDNNOT19	pINCY	Library was constructed using RNA isolated from kidney tissue of a 65-year-old Caucasian male during an exploratory laparotomy and nephroureterectomy. Pathology for the associated tumor tissue indicated a grade 1 renal cell carcinoma within the upper pole of the left kidney. Patient history included malignant melanoma of the abdominal skin, benign neoplasm of the colon, cerebrovascular disease, and umbilical hernia. Family history included myocardial infarction, atherosclerotic coronary artery disease, cerebrovascular disease, prostate cancer, myocardial infarction, and atherosclerotic coronary artery disease.
LATRNOT01	PBLUESCRIPT	Library was constructed using RNA isolated from the left atrium of a 51-year-old Caucasian female, who died from an intracranial bleed.
LATRTUT02	pINCY	Library was constructed using RNA isolated from a myxoma removed from the left atrium of a 43-year-old Caucasian male during annuloplasty. Pathology indicated atrial myxoma. Patient history included pulmonary insufficiency, acute myocardial infarction, atherosclerotic coronary artery disease, hyperlipidemia, and tobacco use. Family history included benign hypertension, acute myocardial infarction, atherosclerotic coronary artery disease, and type II diabetes.

Table 6

Library	Vector	Library Description
LIVRDIT02	pINCY	Library was constructed using RNA isolated from diseased liver tissue removed from a 63-year-old Caucasian female during a liver transplant. Patient history included primary biliary cirrhosis.
LIVRFET02	pINCY	Library was constructed using RNA isolated from liver tissue removed from a Caucasian female fetus, who died at 20 weeks' gestation. Family history included seven days of erythromycin treatment for bronchitis in the mother during the first trimester.
LIVRNOT01	PBLUESCRIPT	Library was constructed at Stratagene, using RNA isolated from the liver tissue of a 49-year-old male.
LIVRTUT04	pINCY	Library was constructed using RNA isolated from liver tumor tissue removed from a 50-year-old Caucasian male during a partial hepatectomy. Pathology indicated a grade 3-4 hepatoma, forming a mass. Patient history included benign hypertension and hepatitis. Hepatitis B core antigen and hepatitis B surface antigen was present in the patient.
LIVRTUT13	pINCY	Library was constructed using RNA isolated from liver tumor tissue removed from a 62-year-old Caucasian female during partial hepatectomy and exploratory laparotomy. Pathology indicated metastatic intermediate grade neuroendocrine carcinoma, consistent with islet cell tumor, forming nodules ranging in size, in the lateral and medial left liver lobe. The pancreas showed fibrosis, chronic inflammation and fat necrosis consistent with pseudocyst. The gall bladder showed mild chronic cholecystitis. Patient history included malignant neoplasm of the pancreas tail, pulmonary embolism, hyperlipidemia, thrombophlebitis, joint pain in multiple joints, type II diabetes, benign hypertension, and cerebrovascular disease. Family history included pancreas cancer, secondary liver cancer, benign hypertension, and hyperlipidemia.
LVENNOT03	PSPORT1	Library was constructed using RNA isolated from the left ventricle tissue of a 31-year-old male.
NERDTDN03	pINCY	This normalized dorsal root ganglion tissue library was constructed from 1.05 million independent clones from a dorsal root ganglion tissue library. Starting RNA was made from dorsal root ganglion tissue removed from the cervical spine of a 32-year-old Caucasian male who died from acute pulmonary edema, acute bronchopneumonia, bilateral pleural effusions, pericardial effusion, and malignant lymphoma (natural killer cell type). The patient presented with pyrexia of unknown origin, malaise, fatigue, and gastrointestinal bleeding. Patient history included probable cytomegalovirus infection, liver congestion, and steatosis, splenomegaly, hemorrhagic cystitis, thyroid hemorrhage, respiratory failure, pneumonia of the left lung, natural killer cell lymphoma of the pharynx, Bell's palsy, and tobacco and alcohol abuse. Previous surgeries included colonoscopy, closed colon biopsy, adenotonsillectomy, and nasopharyngeal endoscopy and biopsy. Patient medications included Diflucan (fluconazole), Deltasone (prednisone),

Table 6

Library	Vector	Library Description
NERDTDN03 - Cont'd		hydrocodone, Lortab, Alprazolam, Reazodone, ProMace-Cytabom, Etoposide, Cisplatin, Cytarabine, and dexamethasone. The patient received radiation therapy and multiple blood transfusions. The library was normalized in 2 rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228-9232 and Bonaldo et al., Genome Research 6 (1996):791, except that a significantly longer (48 hours/round) reannealing hybridization was used.
PHOSDNV44	PCR2-TOPOTA	Library was constructed using pooled cDNA from different donors. cDNA was generated using mRNA isolated from pooled skeletal muscle tissue removed from ten 21 to 57-year-old Caucasian male and female donors who died from sudden death; from pooled thymus tissue removed from nine 18 to 32-year-old Caucasian male and female donors who died from sudden death; from pooled liver tissue removed from 32 Caucasian male and female fetuses who died at 18-24 weeks gestation due to spontaneous abortion; from kidney tissue removed from 59 Caucasian male and female fetuses who died at 20-33 weeks gestation due to spontaneous abortion; and from brain tissue removed from a Caucasian male fetus who died at 23 weeks gestation due to fetal demise.
PITUDIR01	PCDNA2.1	This random primed library was constructed using RNA isolated from pituitary gland tissue removed from a 70-year-old female who died from metastatic adenocarcinoma.
PITUNON01	pINCY	This normalized pituitary gland tissue library was constructed from 6.92 million independent clones from a pituitary gland tissue library. Starting RNA was made from pituitary gland tissue removed from a 55-year-old male who died from chronic obstructive pulmonary disease. Neuropathology indicated there were no gross abnormalities, other than mild ventricular enlargement. There was no apparent microscopic abnormality in any of the neocortical areas examined, except for a number of silver positive neurons with apical dendrite staining, particularly in the frontal lobe. The significance of this was undetermined. The only other microscopic abnormality was that there was prominent silver staining with some swollen axons in the CA3 region of the anterior and posterior hippocampus. Microscopic sections of the cerebellum revealed mild Bergmann's gliosis in the Purkinje cell layer. Patient history included schizophrenia. The library was normalized in two rounds using conditions adapted from Soares et al., PNAS (1994) 91:9228-9232 and Bonaldo et al., Genome Research (1996) 6:791, except that a significantly longer (48 hours/round) reannealing hybridization was used.
PROSBPT02	pINCY	Library was constructed using RNA isolated from diseased prostate tissue removed from a 65-year-old Caucasian male during a radical prostatectomy. Pathology indicated benign prostatic hyperplasia (BPH). Pathology for the associated tumor tissue indicated adenocarcinoma, Gleason grade 3+4. One (of 7) right pelvic lymph nodes was positive for metastatic adenocarcinoma. The patient presented with induration and elevated prostate specific antigen (PSA). Patient history included a benign neoplasm of the large bowel and benign hypertension.

Table 6

Library	Vector	Library Description
PROSTMY01	pINCY	This large size-fractionated cDNA and normalized library was constructed using RNA isolated from diseased prostate tissue removed from a 55-year-old Caucasian male during closed prostatic biopsy, radical prostatectomy, and regional lymph node excision. Pathology indicated adenofibromatous hyperplasia. Pathology for the matched tumor tissue indicated adenocarcinoma Gleason grade 4 forming a predominant mass involving the left side peripherally with extension into the right posterior superior region. The tumor invaded the capsule and perforated the capsule to involve periprostatic tissue in the left posterior superior region. The left inferior posterior and left superior posterior surgical margins are positive. One left pelvic lymph node is metastatically involved. Patient history included calculus of the kidney. Family history included lung cancer and breast cancer. The size-selected library was normalized in 1 round using conditions adapted from Soares et al., PNAS (1994) 91:9228-9232 and Bonaldo et al., Genome Research (1996) 6:791.
PROSTUT12	pINCY	Library was constructed using RNA isolated from prostate tumor tissue removed from a 65-year-old Caucasian male during a radical prostatectomy. Pathology indicated an adenocarcinoma (Gleason grade 2+2). Adenofibromatous hyperplasia was also present. The patient presented with elevated prostate specific antigen (PSA).
PROTDNV28	PCR2-TopoTA	C Unclassified/Mixed Library was constructed using pooled cDNA from different donors. cDNA was generated using mRNA isolated from pooled small intestine tissue removed from a Caucasian male fetus (donor A) who died at 23 weeks' gestation from premature birth; from lung tissue removed from a Caucasian male fetus (donor B) who died from fetal demise; from pleura tumor tissue removed from a 55-year-old Caucasian female (donor C) during a complete pneumonectomy; from frontal/parietal brain tumor tissue removed from a 2-year-old Caucasian female (donor D) during excision of cerebral meningeal lesion; from liver tumor tissue removed from a 72-year-old Caucasian male (donor E) during partial hepatectomy; from pooled fetal brain tissue removed from a Caucasian male fetus (donor F) who was stillborn with a hypoplastic left heart at 23 weeks' gestation and from brain tissue removed from a Caucasian male fetus (donor G) generated using mRNA isolated from pooled small intestine tissue removed from a Caucasian male fetus (donor A) who died at 23 weeks' gestation from premature birth; from lung tissue removed from a

Table 6

Library	Vector	Library Description
PROTDNV28 - Cont'd.		Caucasian male fetus (donor B) who died from fetal demise; from pleura tumor tissue removed from a 55-year-old Caucasian female (donor C) during a complete pneumonectomy; from frontal/parietal brain tumor tissue removed from a 2-year-old Caucasian female (donor D) during excision of cerebral meningeal lesion; from liver tumor tissue removed from a 72-year-old Caucasian male (donor E) during partial hepatectomy; from pooled fetal brain tissue removed from a Caucasian male fetus (donor F) who was stillborn with a hypoplastic left heart at 23 weeks' gestation and from brain tissue removed from a Caucasian male fetus (donor G), who died at 23 weeks' gestation from premature birth; from pooled fetal kidney tissue removed from 59, 20-33-week-old male and female fetuses who died from spontaneous abortion; from pooled thymus tissue removed from 9, 18-32-year-old male and female donors who died from sudden death; and from pooled fetal liver tissue removed from 32, 18-24-week-old male and female fetuses. For donor A, serologies were negative. Family history included diabetes in the mother. For donor B, Serologies were negative. For donor C, pathology indicated grade 3 sarcoma most consistent with leiomyosarcoma, uterine primary, forming a bosellated mass replacing the right lower lobe and a portion of the middle lobe. Multiple nodules comprising the tumor show near total necrosis. Smooth muscle actin was positive. Estrogen receptor was negative and progesterone receptor was positive. The patient presented with shortness of breath. Patient history included peptic ulcer disease, normal delivery, anemia, and tobacco abuse in remission. Previous surgeries included total abdominal hysterectomy, bilateral salpingo-oophorectomy, hemorrhoidectomy, endoscopic excision of lung lesion, and appendectomy. Patient medications included Megace, tamoxifen, and Pepcid. Family history included multiple sclerosis in the mother; atherosclerotic coronary artery disease and type II diabetes in the father; and breast cancer in the grandparent(s). For donor D, pathology indicated neuroectodermal tumor with advanced ganglionic differentiation. The lesion was only moderately cellular but was mitotically active with a high MIB-1 labelling index. Neuronal differentiation was widespread and advanced. Multinucleate and dysplastic-appearing forms were readily seen. The glial element was less prominent. The patient presented with motor seizures. Family history included hypertension in the grandparent(s). For donor E, pathology indicated metastatic grade 2 (of 4) neuroendocrine carcinoma forming a mass. The patient presented with metastatic liver cancer. Patient history included benign hypertension, type I diabetes, prostatic hyperplasia, prostate cancer, alcohol abuse in remission, and tobacco abuse in remission. Previous surgeries included destruction of a pancreatic lesion, closed prostatic biopsy, transurethral prostatectomy, removal of bilateral testes and total splenectomy. Patient medications included Eulexin, Hytrin, Proscar, Ecotrin, and insulin. Family history included atherosclerotic coronary artery disease and acute myocardial infarction in the mother; atherosclerotic coronary artery disease and type II diabetes in the father. For donor F and G, Serologies were negative for both donors and family history for donor G included diabetes in the mother.

Table 6

Library	Vector	Library Description
SINTFER02	pINCY	This random primed library was constructed using RNA isolated from small intestine tissue removed from a Caucasian male fetus who died from fetal demise.
SINTNOR01	PCDNA2.1	This random primed library was constructed using RNA isolated from small intestine tissue removed from a 31-year-old Caucasian female during Roux-en-Y gastric bypass. Patient history included clinical obesity.
SINTNOT19	pINCY	Library was constructed using RNA isolated from small intestine tissue removed from a 8-year-old Black male, who died from anoxia. Serologies were negative. Patient medications included DDAVP, Versed, and labetalol.
SMCCNON03	pINCY	This normalized smooth muscle cell library was constructed from 7.56 million independent clones from a smooth muscle cell library. Starting RNA was made from smooth muscle cell tissue removed from the coronary artery of a 3-year-old Caucasian male. The normalization and hybridization conditions were adapted from Soares et al., (PNAS (1994) 91:9228-9232); Swaroop et al., (NAR (1991) 19:1954); and Bonaldo et al., (Genome Research (1996) 6:791-806), using a significantly longer (48 hour) reannealing hybridization period.
SPLNNOT02	PBLUESCRIPT	Library was constructed using RNA isolated from the spleen tissue of a 29-year-old Caucasian male, who died from head trauma. Serologies were positive for cytomegalovirus (CMV) but otherwise negative. Patient history included alcohol, marijuana, and tobacco use.
SYNORAT03	PSPORT1	Library was constructed using RNA isolated from the wrist synovial membrane tissue of a 56-year-old female with rheumatoid arthritis.
TESTTUT02	pINCY	Library was constructed using RNA isolated from testicular tumor removed from a 31-year-old Caucasian male during unilateral orchiectomy. Pathology indicated embryonal carcinoma.
THP1NOT03	pINCY	Library was constructed using RNA isolated from untreated THP-1 cells. THP-1 is a human promonocyte line derived from the peripheral blood of a 1-year-old Caucasian male with acute monocytic leukemia (ref: Int. J. Cancer (1980) 26:171).
THYMFET02	pINCY	Library was constructed using RNA isolated from thymus tissue removed from a Caucasian female fetus, who died at 17 weeks gestation from anencephalus.

Table 7

Program	Description	Reference	Parameter Threshold
ABI FACTURA	A program that removes vector sequences and masks ambiguous bases in nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
ABI/PARACEL FDF	A Fast Data Finder useful in comparing and annotating amino acid or nucleic acid sequences.	Applied Biosystems, Foster City, CA; Paracel Inc., Pasadena, CA.	Mismatch <50%
ABI AutoAssembler	A program that assembles nucleic acid sequences.	Applied Biosystems, Foster City, CA.	
BLAST	A Basic Local Alignment Search Tool useful in sequence similarity search for amino acid and nucleic acid sequences. BLAST includes five functions: blastp, blastn, blastx, tblastn, and tblastx.	Altschul, S.F. et al. (1990) J. Mol. Biol. 215:403-410; Altschul, S.F. et al. (1997) Nucleic Acids Res. 25:3389-3402.	ESTs: Probability value = 1.0E-8 or less; Full Length sequences: Probability value = 1.0E-10 or less
FASTA	A Pearson and Lipman algorithm that searches for similarity between a query sequence and a group of sequences of the same type. FASTA comprises as least five functions: fasta, tfasta, fastx, tfastx, and ssearch.	Pearson, W.R. and D.J. Lipman (1988) Proc. Natl. Acad Sci. USA 85:2444-2448; Pearson, W.R. (1990) Methods Enzymol. 183:63-98; and Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489.	ESTs: fasta E value = 1.06E-6; Assembled ESTs: fasta Identity = 95% or greater and Match length = 200 bases or greater; fastx E value = 1.0E-8 or less; Full Length sequences: fastx score = 100 or greater
BLIMPS	A BLocks IMProved Searcher that matches a sequence against those in BLOCKS, PRINTS, DOMO, PRODOM, and PFAM databases to search for gene families, sequence homology, and structural fingerprint regions.	Henikoff, S. and J.G. Henikoff (1991) Nucleic Acids Res. 19:6565-6572; Henikoff, J.G. and S. Henikoff (1996) Methods Enzymol. 266:88-105; and Attwood, T.K. et al. (1997) J. Chem. Inf. Comput. Sci. 37:417-424.	Probability value = 1.0E-3 or less

Table 7

Program	Description	Reference	Parameter Threshold
HMMER	An algorithm for searching a query sequence against hidden Markov model (HMM)-based databases of protein family consensus sequences, such as PFAM, INCY, SMART and TIGRFAM.	Krogh, A. et al. (1994) J. Mol. Biol. 235:1501-1531; Sonnhammer, E.L.L. et al. (1988) Nucleic Acids Res. 26:320-322; Durbin, R. et al. (1998) Our World View, in a Nutshell, Cambridge Univ. Press, pp. 1-350.	PFAM, INCY, SMART or TIGRFAM hits: Probability value = 1.0E-3 or less; Signal peptide hits: Score = 0 or greater
ProfileScan	An algorithm that searches for structural and sequence motifs in protein sequences that match sequence patterns defined in Prosite.	Gribskov, M. et al. (1988) CABIOS 4:61-66; Gribskov, M. et al. (1989) Methods Enzymol. 183:146-159; Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221.	Normalized quality score ≥ GCG specified "HIGH" value for that particular Prosite motif. Generally, score = 1.4-2.1.
Phred	A base-calling algorithm that examines automated sequencer traces with high sensitivity and probability.	Ewing, B. et al. (1998) Genome Res. 8:175-185; Ewing, B. and P. Green (1998) Genome Res. 8:186-194.	
Phrap	A Phils Revised Assembly Program including SWAT and CrossMatch, programs based on efficient implementation of the Smith-Waterman algorithm, useful in searching sequence homology and assembling DNA sequences.	Smith, T.F. and M.S. Waterman (1981) Adv. Appl. Math. 2:482-489; Smith, T.F. and M.S. Waterman (1981) J. Mol. Biol. 147:195-197; and Green, P., University of Washington, Seattle, WA.	Score = 120 or greater; Match length = 56 or greater
Consed	A graphical tool for viewing and editing Phrap assemblies.	Gordon, D. et al. (1998) Genome Res. 8:195-202.	
SPScan	A weight matrix analysis program that scans protein sequences for the presence of secretory signal peptides.	Nielson, H. et al. (1997) Protein Engineering 10:1-6; Claverie, J.M. and S. Audic (1997) CABIOS 12:431-439.	Score = 3.5 or greater
TMAP	A program that uses weight matrices to delineate transmembrane segments on protein sequences and determine orientation.	Persson, B. and P. Argos (1994) J. Mol. Biol. 237:182-192; Persson, B. and P. Argos (1996) Protein Sci. 5:363-371.	

Table 7

Program	Description	Reference	Parameter Threshold
TMHMMER	A program that uses a hidden Markov model (HMM) to delineate transmembrane segments on protein sequences and determine orientation.	Sonnhammer, E.L. et al. (1998) Proc. Sixth Intl. Conf. On Intelligent Systems for Mol. Biol., Glasgow et al., eds., The Am. Assoc. for Artificial Intelligence (AAAI) Press, Menlo Park, CA, and MIT Press, Cambridge, MA, pp. 175-182.	
Motifs	A program that searches amino acid sequences for patterns that matched those defined in Prosite.	Bairoch, A. et al. (1997) Nucleic Acids Res. 25:217-221; Wisconsin Package Program Manual, version 9, page M51-59, Genetics Computer Group, Madison, WI.	

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
56	7509350	1347961H1	SNP00140369	54	356	T	T	C	M116	n/a	n/a	n/a	n/a
56	7509350	1347961H1	SNP00140370	215	517	A	A	G	I170	n/a	n/a	n/a	n/a
56	7509350	1805095H1	SNP00015376	163	157	C	C	T	L50	n/a	n/a	n/a	n/a
56	7509350	1805095H1	SNP00140368	127	121	C	C	A	R38	n/a	n/a	n/a	n/a
56	7509350	3037428H1	SNP00140368	144	119	C	C	A	T37	n/a	n/a	n/a	n/a
56	7509350	3213432H1	SNP00015376	161	152	C	C	T	S48	n/a	n/a	n/a	n/a
56	7509350	3213432H1	SNP00140368	125	116	C	C	A	T36	n/a	n/a	n/a	n/a
56	7509350	3254661H1	SNP00015376	176	158	C	C	T	P50	n/a	n/a	n/a	n/a
56	7509350	3254661H1	SNP00140368	140	122	C	C	A	P38	n/a	n/a	n/a	n/a
56	7509350	3405543H1	SNP00140370	152	515	A	A	G	E169	n/a	n/a	n/a	n/a
56	7509350	3512549H1	SNP00015376	141	156	T	C	T	R49	n/a	n/a	n/a	n/a
56	7509350	3512549H1	SNP00140368	105	120	C	C	A	S37	n/a	n/a	n/a	n/a
56	7509350	3577339H1	SNP00015376	155	154	C	C	T	R49	n/a	n/a	n/a	n/a
56	7509350	3577339H1	SNP00140368	119	118	C	C	A	R37	n/a	n/a	n/a	n/a
56	7509350	3676454H1	SNP00140368	125	117	C	C	A	N36	n/a	n/a	n/a	n/a
56	7509350	3944013H1	SNP00015376	155	155	C	C	T	P49	n/a	n/a	n/a	n/a
56	7509350	4171472H1	SNP00015376	155	153	C	C	T	C48	n/a	n/a	n/a	n/a
56	7509350	4641077H1	SNP00140369	152	355	T	T	C	L116	n/a	n/a	n/a	n/a
56	7509350	4834894H1	SNP00140371	165	1537	A	A	G	noncoding	n/a	n/a	n/a	n/a
56	7509350	4834894H1	SNP00140372	173	1545	T	T	C	noncoding	n/a	n/a	n/a	n/a
56	7509350	4876888H1	SNP00140369	30	354	T	T	C	P115	n/a	n/a	n/a	n/a
56	7509350	6495896H1	SNP00015377	342	1497	C	C	T	noncoding	n/a	n/a	n/a	n/a
56	7509350	7748086H1	SNP00140371	92	1538	A	A	G	noncoding	n/a	n/a	n/a	n/a
56	7509350	7748086H1	SNP00140372	100	1546	T	T	C	noncoding	n/a	n/a	n/a	n/a
57	7509325	086872H1	SNP00073588	43	1450	G	A	G	noncoding	n/a	n/a	n/a	n/a
57	7509325	4419451H1	SNP00027681	137	683	C	C	T	noncoding	n/a	n/a	n/a	n/a
57	7509325	5396974H1	SNP00043998	221	1751	C	C	T	noncoding	n/a	n/a	n/a	n/a
57	7509325	6895504H1	SNP00110389	499	670	G	G	A	noncoding	0.97	n/d	n/d	n/d

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
57	7509325	6899923H1	SNP00110389	476	663	G	G	A	noncoding	0.97	n/d	n/d	n/d
58	7509337	3981508H1	SNP00066709	204	1123	C	C	T	noncoding	n/d	n/a	n/a	n/a
59	7509353	2130414H1	SNP00131705	60	1842	C	C	A	noncoding	n/a	n/a	n/a	n/a
59	7509353	2512629H1	SNP00013873	51	88	G	G	A	noncoding	0.28	0.62	0.13	0.28
59	7509353	2514562H1	SNP00013873	99	90	G	G	A	noncoding	0.28	0.62	0.13	0.28
59	7509353	2876718H1	SNP00131705	227	1840	C	C	A	noncoding	n/a	n/a	n/a	n/a
59	7509353	3503573H1	SNP00131705	220	1839	C	C	A	noncoding	n/a	n/a	n/a	n/a
62	7509216	1917632H1	SNP00127501	207	1241	C	C	A	noncoding	n/a	n/a	n/a	n/a
62	7509216	281088H1	SNP00065806	81	502	C	T	C	S28	n/a	n/a	n/a	n/a
62	7509216	6347325H1	SNP00127501	56	1239	C	C	A	noncoding	n/a	n/a	n/a	n/a
62	7509216	7000493H1	SNP00065806	369	520	T	T	C	G34	n/a	n/a	n/a	n/a
63	7509376	1917632H1	SNP00127501	207	1279	C	C	A	noncoding	n/a	n/a	n/a	n/a
63	7509376	281088H1	SNP00065806	81	498	C	T	C	S28	n/a	n/a	n/a	n/a
63	7509376	6347325H1	SNP00127501	56	1277	C	C	A	noncoding	n/a	n/a	n/a	n/a
63	7509376	7000493H1	SNP00065806	369	516	T	T	C	G34	n/a	n/a	n/a	n/a
64	7501927	1388825H1	SNP00023309	42	599	T	T	C	C188	0.51	0.45	0.32	0.38
64	7501927	2224334H1	SNP00039353	60	1313	C	C	A	noncoding	n/a	n/a	n/a	n/a
64	7501927	2681643H1	SNP00039351	56	71	T	T	C	stop12	n/a	n/a	n/a	n/a
64	7501927	3040488H1	SNP00039351	82	84	C	T	C	T16	n/a	n/a	n/a	n/a
64	7501927	3040488H1	SNP00039352	261	263	A	G	A	N76	n/a	n/a	n/a	n/a
64	7501927	3244023H1	SNP00147759	34	488	T	T	C	L151	n/a	n/a	n/a	n/a
64	7501927	3244023H1	SNP00147760	61	515	T	T	C	L160	n/a	n/a	n/a	n/a
64	7501927	3533453H1	SNP00120338	285	1067	T	C	T	F344	0.64	0.80	0.77	0.77
64	7501927	3725331H1	SNP00039351	81	79	T	T	C	V14	n/a	n/a	n/a	n/a
64	7501927	4068886H1	SNP00039353	214	1306	C	C	A	noncoding	n/a	n/a	n/a	n/a
64	7501927	4793930H1	SNP00039351	79	83	T	T	C	L16	n/a	n/a	n/a	n/a
64	7501927	4793930H1	SNP00039352	258	262	G	G	A	K75	n/a	n/a	n/a	n/a
64	7501927	6821531H1	SNP00120338	136	1066	T	C	T	H343	0.64	0.80	0.77	0.77

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
64	7501927	7403716H1	SNP00120339	40	1166	C	C	T	Q377	n/a	n/a	n/a	n/a
64	7501927	7634971J1	SNP00120340	71	1234	C	C	T	A399	n/a	n/a	n/a	n/a
64	7501927	7640189H1	SNP00133429	389	875	G	G	A	G280	n/a	n/a	n/a	n/a
65	7503274	1354256H1	SNP00061150	216	591	G	G	T	Q191	n/d	n/d	n/d	n/a
65	7503274	1363539H1	SNP00110597	126	1217	T	T	C	V400	n/d	n/a	n/a	n/a
65	7503274	1548963H1	SNP00061152	18	1058	C	C	T	A347	n/a	n/a	n/a	n/a
65	7503274	3336111H1	SNP00061150	230	589	G	G	T	E191	n/d	n/d	n/d	n/a
65	7503274	3406517H1	SNP00110597	25	1215	T	T	C	V399	n/d	n/a	n/a	n/a
65	7503274	3558394H1	SNP00061152	246	1057	C	C	T	P347	n/a	n/a	n/a	n/a
65	7503274	3570601H1	SNP00110597	95	1216	T	T	C	F400	n/d	n/a	n/a	n/a
65	7503274	3845993H1	SNP00110597	155	1214	T	T	C	V399	n/d	n/a	n/a	n/a
65	7503274	3873385H1	SNP00061150	4	588	G	G	T	E190	n/d	n/d	n/d	n/a
65	7503274	4643668H1	SNP00061150	59	582	G	G	T	R188	n/d	n/d	n/d	n/a
65	7503274	4743084H1	SNP00061152	181	1056	C	C	T	C346	n/a	n/a	n/a	n/a
65	7503274	5711037H1	SNP00061150	49	590	G	G	T	R191	n/d	n/d	n/d	n/a
65	7503274	6309978H1	SNP00061152	267	1053	C	C	T	T345	n/a	n/a	n/a	n/a
65	7503274	6309978H1	SNP00110597	426	1212	T	T	C	P398	n/d	n/a	n/a	n/a
66	7509104	6788468H1	SNP00098196	397	895	T	C	T	M278	n/d	n/d	n/d	n/d
66	7509104	6788468H1	SNP00132112	305	806	C	C	T	Y248	n/a	n/a	n/a	n/a
66	7509104	7636931H1	SNP00098196	330	550	C	C	T	P163	n/d	n/d	n/d	n/d
67	7509996	099563H1	SNP00029752	111	1573	A	A	G	noncoding	n/a	n/a	n/a	n/a
67	7509996	099563H1	SNP00101027	55	1517	A	A	G	noncoding	n/a	n/a	n/a	n/a
67	7509996	1379082H1	SNP00029752	153	1584	A	A	G	noncoding	n/a	n/a	n/a	n/a
67	7509996	1379082H1	SNP00101027	97	1528	A	A	G	noncoding	n/a	n/a	n/a	n/a
67	7509996	1417733H1	SNP00132940	94	411	C	C	T	F114	n/a	n/a	n/a	n/a
67	7509996	1417737H1	SNP00029752	109	1585	A	A	G	noncoding	n/a	n/a	n/a	n/a
67	7509996	1417737H1	SNP00101027	53	1529	A	A	G	noncoding	n/a	n/a	n/a	n/a
67	7509996	1417910H1	SNP00101026	15	1370	A	A	G	noncoding	n/a	n/a	n/a	n/a

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
67	7509996	1418262H1	SNP00101024	40	539	T	T	C	L157	n/d	n/d	n/d	n/d
67	7509996	1418992H1	SNP00101025	110	917	A	A	G	D283	n/d	n/d	n/d	n/d
67	7509996	1596309H1	SNP00094855	63	1113	G	C	G	L348	0.51	0.68	0.44	0.55
67	7509996	2365849H1	SNP00029752	218	1587	A	A	G	noncoding	n/a	n/a	n/a	n/a
67	7509996	2367556H1	SNP00029751	7	36	T	T	C	noncoding	n/a	n/a	n/a	n/a
67	7509996	2371091H1	SNP00101023	129	370	A	A	G	T101	n/a	n/a	n/a	n/a
67	7509996	2562751H1	SNP00101024	150	540	T	T	C	L157	n/d	n/d	n/d	n/d
67	7509996	2664571H1	SNP00029752	163	1583	A	A	G	noncoding	n/a	n/a	n/a	n/a
67	7509996	2664571H1	SNP00101027	107	1527	A	A	G	noncoding	n/a	n/a	n/a	n/a
67	7509996	2961063H1	SNP00029751	6	35	T	T	C	noncoding	n/a	n/a	n/a	n/a
67	7509996	3501625H1	SNP00101025	287	914	A	A	G	Q282	n/d	n/d	n/d	n/d
67	7509996	3501633H1	SNP00101024	166	537	T	T	C	I156	n/d	n/d	n/d	n/d
67	7509996	3501633H1	SNP00132940	38	409	C	C	T	L114	n/a	n/a	n/a	n/a
67	7509996	3503180H1	SNP00101023	31	365	A	A	G	K99	n/a	n/a	n/a	n/a
67	7509996	3503710H1	SNP00101026	179	1368	A	A	G	noncoding	n/a	n/a	n/a	n/a
67	7509996	3504402H1	SNP00101026	178	1369	A	A	G	noncoding	n/a	n/a	n/a	n/a
67	7509996	3506203H1	SNP00101023	63	369	A	A	G	A100	n/a	n/a	n/a	n/a
67	7509996	6812405J1	SNP00101022	87	120	G	G	A	W17	n/d	n/d	n/d	n/d
67	7509996	6812405J1	SNP00132939	42	76	G	G	A	E3	n/a	n/a	n/a	n/a
67	7509996	7650236H2	SNP00006435	156	207	C	C	T	H46	0.47	n/a	n/a	n/a
67	7509996	7650236H2	SNP00006436	99	264	G	G	T	S65	n/a	n/a	n/a	n/a
67	7509996	941474H1	SNP00101026	285	1378	A	A	G	noncoding	n/a	n/a	n/a	n/a
68	7510030	182066H1	SNP00010549	111	633	T	T	C	I197	0.46	0.45	0.33	0.35
68	7510030	3530258H1	SNP00040813	107	942	G	G	A	noncoding	n/a	n/a	n/a	n/a
68	7510030	3530258H1	SNP00099284	99	934	T	T	C	noncoding	n/a	n/a	n/a	n/a
68	7510030	3593785H1	SNP00050275	69	108	A	A	G	Q22	n/a	n/a	n/a	n/a
68	7510030	4539814H1	SNP00040813	93	945	G	G	A	noncoding	n/a	n/a	n/a	n/a
68	7510030	4539814H1	SNP00099284	85	937	T	T	C	noncoding	n/a	n/a	n/a	n/a

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
68	7510030	4590421H1	SNP00010549	100	706	C	T	C	A221	0.46	0.45	0.33	0.35
68	7510030	5562718H1	SNP00010549	218	635	C	T	C	P198	0.46	0.45	0.33	0.35
68	7510030	6834341H1	SNP00010549	270	709	C	T	C	A222	0.46	0.45	0.33	0.35
69	7510062	1560250H1	SNP00107470	86	2138	A	A	G	noncoding	0.96	n/a	n/a	n/a
69	7510062	1784861H1	SNP00009615	192	2968	T	T	G	noncoding	n/a	n/a	n/a	n/a
69	7510062	2196685H1	SNP00037844	78	1908	G	G	T	noncoding	n/d	n/a	n/a	n/a
69	7510062	2350164H1	SNP00009614	99	3021	C	C	T	noncoding	0.99	0.97	0.97	n/a
69	7510062	2439943H1	SNP00107471	186	2288	G	G	A	noncoding	n/d	n/d	n/d	n/d
69	7510062	2554655H1	SNP00052935	166	856	C	C	G	A206	n/d	n/d	n/d	n/d
69	7510062	3112878H1	SNP00037844	31	1906	G	G	T	noncoding	n/d	n/a	n/a	n/a
69	7510062	3112878H1	SNP00107470	260	2136	A	A	G	noncoding	0.96	n/a	n/a	n/a
69	7510062	3369632H1	SNP00107470	61	2137	A	A	G	noncoding	0.96	n/a	n/a	n/a
69	7510062	3369632H1	SNP00107471	211	2287	G	G	A	noncoding	n/d	n/d	n/d	n/d
69	7510062	3508437H1	SNP00009614	208	3018	C	C	T	noncoding	0.99	0.97	0.97	n/a
69	7510062	3508437H1	SNP00009615	155	2965	T	T	G	noncoding	n/a	n/a	n/a	n/a
69	7510062	3523022H1	SNP00009614	252	3020	C	C	T	noncoding	0.99	0.97	0.97	n/a
69	7510062	3523022H1	SNP00009615	199	2967	T	T	G	noncoding	n/a	n/a	n/a	n/a
69	7510062	3679504H1	SNP00107471	110	2286	G	G	A	noncoding	n/d	n/d	n/d	n/d
69	7510062	3875558H1	SNP00009614	181	3019	C	C	T	noncoding	0.99	0.97	0.97	n/a
69	7510062	3875558H1	SNP00009615	128	2966	T	T	G	noncoding	n/a	n/a	n/a	n/a
69	7510062	6341477H1	SNP00120597	416	3157	A	G	A	noncoding	n/a	n/a	n/a	n/a
70	7510217	1390802H1	SNP00014933	81	1090	G	G	A	noncoding	n/d	n/a	n/a	n/a
70	7510217	2241230H1	SNP00041758	27	1153	A	A	G	noncoding	n/a	n/a	n/a	n/a
70	7510217	2572117H1	SNP00041757	182	705	T	C	T	noncoding	n/a	n/a	n/a	n/a
70	7510217	3249891H1	SNP00014932	112	446	C	C	A	G94	n/a	n/a	n/a	n/a
70	7510217	3406412H1	SNP00041758	141	1152	A	A	G	noncoding	n/a	n/a	n/a	n/a
70	7510217	4410894H1	SNP00041758	72	1150	A	A	G	noncoding	n/a	n/a	n/a	n/a
70	7510217	445875H1	SNP00041757	97	602	T	C	T	F146	n/a	n/a	n/a	n/a

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
70	7510217	6055006H1	SNP00106542	132	449	G	G	C	R95	n/a	n/a	n/a	n/a
70	7510217	6749882H1	SNP00041757	315	605	T	C	T	S147	n/a	n/a	n/a	n/a
70	7510217	6996758H1	SNP00041758	284	1026	A	A	G	noncoding	n/a	n/a	n/a	n/a
70	7510217	773479H1	SNP00014933	64	1083	G	G	A	noncoding	n/d	n/a	n/a	n/a
71	7510298	027097H1	SNP00131732	201	1630	A	A	G	noncoding	n/a	n/a	n/a	n/a
71	7510298	1319852H1	SNP00036036	33	1224	G	G	A	noncoding	n/a	n/a	n/a	n/a
71	7510298	1319852H1	SNP00121085	201	1392	A	A	G	noncoding	n/d	n/d	n/d	n/d
71	7510298	1343264H1	SNP00092568	188	1898	C	C	G	noncoding	n/a	n/a	n/a	n/a
71	7510298	2092278H1	SNP00114196	176	2078	T	T	C	noncoding	n/d	n/d	n/d	n/d
71	7510298	2869345H1	SNP00121085	9	1386	A	A	G	noncoding	n/d	n/d	n/d	n/d
71	7510298	292088H1	SNP00114196	72	2079	T	T	C	noncoding	n/d	n/d	n/d	n/d
71	7510298	3460479H1	SNP00131732	200	1627	A	A	G	noncoding	n/a	n/a	n/a	n/a
71	7510298	3769630H1	SNP00121085	104	1390	A	A	G	noncoding	n/d	n/d	n/d	n/d
71	7510298	3804827H1	SNP00121085	186	1387	A	A	G	noncoding	n/d	n/d	n/d	n/d
71	7510298	3871190H1	SNP00092568	71	1905	C	C	G	noncoding	n/a	n/a	n/a	n/a
71	7510298	3871190H1	SNP00114196	252	2086	T	T	C	noncoding	n/d	n/d	n/d	n/d
71	7510298	3997022H1	SNP00114196	154	2077	T	T	C	noncoding	n/d	n/d	n/d	n/d
71	7510298	4370665H1	SNP00114196	120	2076	T	T	C	noncoding	n/d	n/d	n/d	n/d
71	7510298	4589941H1	SNP00121085	55	1388	A	A	G	noncoding	n/d	n/d	n/d	n/d
71	7510298	4856203H1	SNP00131732	70	1628	A	A	G	noncoding	n/a	n/a	n/a	n/a
71	7510298	5064517H1	SNP00114195	39	1347	G	G	C	noncoding	n/a	n/a	n/a	n/a
71	7510298	5099688H1	SNP00131732	29	1629	A	A	G	noncoding	n/a	n/a	n/a	n/a
71	7510298	5106110H1	SNP00092568	199	1897	C	C	G	noncoding	n/a	n/a	n/a	n/a
71	7510298	5492414H1	SNP00131732	102	1631	A	A	G	noncoding	n/a	n/a	n/a	n/a
71	7510298	5787670H1	SNP00114195	87	1345	G	G	C	noncoding	n/a	n/a	n/a	n/a
71	7510298	5809384H1	SNP00092568	117	1896	C	C	G	noncoding	n/a	n/a	n/a	n/a
71	7510298	5858785H1	SNP00131732	81	1625	A	A	G	noncoding	n/a	n/a	n/a	n/a
71	7510298	6128484H1	SNP00066861	306	1761	T	C	T	noncoding	n/d	n/d	n/d	n/d

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
71	7510298	6350191H1	SNP00092568	136	1895	C	C	G	noncoding	n/a	n/a	n/a	n/a
71	7510298	6828961J1	SNP00008648	539	768	G	G	C	K168	n/d	n/d	n/d	n/d
71	7510298	875690H1	SNP00121085	147	1391	A	A	G	noncoding	n/d	n/d	n/d	n/d
72	7510299	1241454H1	SNP00108587	139	404	C	C	T	P118	n/a	n/a	n/a	n/a
72	7510299	1807704H1	SNP00068801	56	225	C	C	G	L58	n/a	n/a	n/a	n/a
72	7510299	2830813H1	SNP00068802	145	445	G	G	C	A132	n/d	n/a	n/d	0.97
72	7510299	2830813H1	SNP00108587	106	406	C	C	T	L119	n/a	n/a	n/a	n/a
72	7510299	2848960H1	SNP00068801	86	226	C	C	G	Q59	n/a	n/a	n/a	n/a
72	7510299	3326981H1	SNP00068801	224	224	C	C	G	P58	n/a	n/a	n/a	n/a
72	7510299	6826290H1	SNP00068801	186	178	C	C	G	L43	n/a	n/a	n/a	n/a
73	7510368	2257046H1	SNP00059656	13	912	T	T	C	noncoding	n/a	n/a	n/a	n/a
73	7510368	2615823H1	SNP00028270	6	36	G	G	T	noncoding	0.98	n/a	n/a	n/a
73	7510368	2615823H1	SNP00028271	29	59	C	C	G	noncoding	n/a	n/a	n/a	n/a
73	7510368	2874677H1	SNP00028271	9	51	C	C	G	noncoding	n/a	n/a	n/a	n/a
73	7510368	3461718H1	SNP00028270	9	34	T	G	T	noncoding	0.98	n/a	n/a	n/a
73	7510368	3462319H1	SNP00059656	296	908	T	T	C	noncoding	n/a	n/a	n/a	n/a
73	7510368	4161369H1	SNP00059656	117	909	T	T	C	noncoding	n/a	n/a	n/a	n/a
73	7510368	4782374H1	SNP00059656	56	887	T	T	C	noncoding	n/a	n/a	n/a	n/a
73	7510368	5840934H2	SNP00028270	44	30	G	G	T	noncoding	0.98	n/a	n/a	n/a
73	7510368	5840934H2	SNP00028271	67	53	C	C	G	noncoding	n/a	n/a	n/a	n/a
73	7510368	5919021H1	SNP00028270	118	33	T	G	T	noncoding	0.98	n/a	n/a	n/a
73	7510368	5970391H1	SNP00059657	17	1012	G	T	G	noncoding	n/a	n/a	n/a	n/a
73	7510368	5970391H1	SNP00059658	455	1450	C	A	C	noncoding	n/a	n/a	n/a	n/a
73	7510368	6413064H1	SNP00059656	38	911	C	T	C	noncoding	n/a	n/a	n/a	n/a
74	7510369	2257046H1	SNP00059656	13	792	T	T	C	I197	n/a	n/a	n/a	n/a
74	7510369	2615823H1	SNP00028270	6	52	G	G	T	noncoding	0.98	n/a	n/a	n/a
74	7510369	2615823H1	SNP00028271	29	75	C	C	G	noncoding	n/a	n/a	n/a	n/a
74	7510369	2874677H1	SNP00028271	9	67	C	C	G	noncoding	n/a	n/a	n/a	n/a

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
74	7510369	3169160H1	SNP00028270	93	51	G	G	T	noncoding	0.98	n/a	n/a	n/a
74	7510369	3461718H1	SNP00028270	9	50	T	G	T	noncoding	0.98	n/a	n/a	n/a
74	7510369	3462319H1	SNP00059656	296	788	T	T	C	S196	n/a	n/a	n/a	n/a
74	7510369	4161369H1	SNP00059656	117	789	T	T	C	F196	n/a	n/a	n/a	n/a
74	7510369	4782374H1	SNP00059656	56	767	T	T	C	stop189	n/a	n/a	n/a	n/a
74	7510369	5840934H2	SNP00028270	44	46	G	G	T	noncoding	0.98	n/a	n/a	n/a
74	7510369	5840934H2	SNP00028271	67	69	C	C	G	noncoding	n/a	n/a	n/a	n/a
74	7510369	5919021H1	SNP00028270	118	49	T	G	T	noncoding	0.98	n/a	n/a	n/a
74	7510369	5970391H1	SNP00059658	455	1430	C	A	C	P410	n/a	n/a	n/a	n/a
74	7510369	6413064H1	SNP00059656	38	791	C	T	C	L197	n/a	n/a	n/a	n/a
74	7510369	6839273H1	SNP00059657	232	892	T	T	G	R230	n/a	n/a	n/a	n/a
75	7510377	2098152H1	SNP00098532	17	1821	T	T	C	noncoding	0.61	0.81	0.77	0.65
75	7510377	2421124H1	SNP00023191	56	1900	C	G	C	noncoding	n/a	n/a	n/a	n/a
75	7510377	2652175H1	SNP00005185	99	2411	G	G	A	noncoding	0.98	n/a	n/a	n/a
75	7510377	6058644H1	SNP00023191	336	1848	G	G	C	noncoding	n/a	n/a	n/a	n/a
75	7510377	6534732H1	SNP00003472	75	1730	C	C	T	noncoding	n/a	n/a	n/a	n/a
75	7510377	7127569H1	SNP00005185	20	2401	G	G	A	noncoding	0.98	n/a	n/a	n/a
75	7510377	7997342H1	SNP00142702	160	600	A	A	G	G111	n/a	n/a	n/a	n/a
76	7510026	1913184H1	SNP00024027	231	4013	G	G	A	noncoding	0.92	0.96	0.95	n/d
76	7510026	2917642H1	SNP00059634	65	3544	A	A	G	noncoding	n/d	0.93	0.76	0.91
76	7510026	3250819H1	SNP00057803	81	2271	T	T	C	S707	n/a	n/a	n/a	n/a
76	7510026	3359092H1	SNP00057803	28	2270	T	T	C	I707	n/a	n/a	n/a	n/a
76	7510026	3719593H1	SNP00024027	89	4008	G	G	A	noncoding	0.92	0.96	0.95	n/d
76	7510026	3767582H1	SNP00024027	20	4012	G	G	A	noncoding	0.92	0.96	0.95	n/d
76	7510026	4247584H1	SNP00059634	120	3546	A	A	G	noncoding	n/d	0.93	0.76	0.91
76	7510026	4319238H1	SNP00057803	199	2268	T	T	C	T706	n/a	n/a	n/a	n/a
76	7510026	4724435H1	SNP00024027	181	3998	A	G	A	noncoding	0.92	0.96	0.95	n/d
76	7510026	5768362H1	SNP00059634	313	3547	G	A	G	noncoding	n/d	0.93	0.76	0.91

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
76	7510026	5810649H1	SNP00057803	22	2269	T	T	C	C707	n/a	n/a	n/a	n/a
77	7509168	1816249H1	SNP00011194	194	386	A	A	G	noncoding	n/a	n/a	n/a	n/a
77	7509168	2091222H1	SNP00001967	151	346	G	C	G	noncoding	0.61	0.41	0.79	0.79
77	7509168	2091222H1	SNP00019502	144	339	C	C	T	noncoding	n/a	n/a	n/a	n/a
77	7509168	2097659H1	SNP00019503	143	640	A	A	C	noncoding	n/a	n/a	n/a	n/a
77	7509168	2097659H1	SNP00052136	262	759	T	T	C	noncoding	n/d	n/a	n/a	n/a
77	7509168	275129H1	SNP00019503	203	641	A	A	C	noncoding	n/a	n/a	n/a	n/a
77	7509168	2961392H1	SNP00011194	123	380	G	A	G	noncoding	n/a	n/a	n/a	n/a
77	7509168	3567516H1	SNP00001967	182	345	C	C	G	noncoding	0.61	0.41	0.79	0.79
77	7509168	3697334H1	SNP00011194	67	384	G	A	G	noncoding	n/a	n/a	n/a	n/a
77	7509168	3925739H1	SNP00001967	227	339	C	C	G	noncoding	0.61	0.41	0.79	0.79
77	7509168	3925739H1	SNP00019502	220	332	C	C	T	noncoding	n/a	n/a	n/a	n/a
77	7509168	4213146H1	SNP00001967	64	343	G	C	G	noncoding	0.61	0.41	0.79	0.79
77	7509168	4213272H1	SNP00019502	57	336	T	C	T	noncoding	n/a	n/a	n/a	n/a
77	7509168	4630177H1	SNP00019503	249	636	A	A	C	noncoding	n/a	n/a	n/a	n/a
77	7509168	4634491H1	SNP00019503	176	637	A	A	C	noncoding	n/a	n/a	n/a	n/a
78	7500607	1439616H1	SNP00132549	28	57	T	T	C	noncoding	n/a	n/a	n/a	n/a
78	7500607	1701044H1	SNP00132550	6	290	G	G	C	noncoding	0.92	n/a	n/a	n/a
78	7500607	1965668H1	SNP00058642	269	2171	C	C	A	noncoding	n/a	n/a	n/a	n/a
78	7500607	2132456H1	SNP00043794	126	2428	C	C	A	noncoding	n/a	n/a	n/a	n/a
78	7500607	284438H1	SNP00000434	143	2941	C	T	C	noncoding	n/a	n/a	n/a	n/a
78	7500607	284438H1	SNP00108726	174	2972	G	G	T	noncoding	n/a	n/a	n/a	n/a
78	7500607	3567092H1	SNP00132549	93	55	T	T	C	noncoding	n/a	n/a	n/a	n/a
78	7500607	3707070H1	SNP00132549	80	56	T	T	C	noncoding	n/a	n/a	n/a	n/a
78	7500607	3867578H1	SNP00132549	43	54	T	T	C	noncoding	n/a	n/a	n/a	n/a
78	7500607	3867578H1	SNP00132550	276	287	G	G	C	noncoding	0.92	n/a	n/a	n/a
78	7500607	3875094H1	SNP00044177	209	2235	A	A	C	noncoding	n/d	n/a	n/a	n/a
78	7500607	3875094H1	SNP00058642	143	2169	C	C	A	noncoding	n/a	n/a	n/a	n/a

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
78	7500607	4593872H1	SNP00043794	115	2425	C	C	A	noncoding	n/a	n/a	n/a	n/a
78	7500607	4779825H1	SNP00044177	124	2237	C	A	C	noncoding	n/d	n/a	n/a	n/a
78	7500607	4994707H1	SNP00132549	170	53	T	T	C	noncoding	n/a	n/a	n/a	n/a
78	7500607	5840196H1	SNP00132549	203	47	T	T	C	noncoding	n/a	n/a	n/a	n/a
78	7500607	6305967H1	SNP00132549	219	51	T	T	C	noncoding	n/a	n/a	n/a	n/a
78	7500607	6305967H1	SNP00132550	452	284	G	G	C	noncoding	0.92	n/a	n/a	n/a
78	7500607	6446668H1	SNP00094323	106	976	C	C	G	noncoding	n/a	n/a	n/a	n/a
78	7500607	6463809H2	SNP00132549	81	23	T	T	C	noncoding	n/a	n/a	n/a	n/a
78	7500607	6814993J1	SNP00112479	56	665	A	A	G	I66	0.95	0.87	0.59	0.79
78	7500607	6834890H1	SNP00000435	39	2977	G	G	A	noncoding	n/a	n/a	n/a	n/a
78	7500607	8664714H1	SNP00043794	551	2420	C	C	A	noncoding	n/a	n/a	n/a	n/a
79	7506079	1336549H1	SNP00024403	18	217	C	C	T	N57	n/a	n/a	n/a	n/a
79	7506079	1421173H1	SNP00024404	75	1393	G	G	C	noncoding	n/a	n/a	n/a	n/a
79	7506079	1572030H1	SNP00053025	161	810	C	C	A	A255	0.96	n/d	n/d	n/d
79	7506079	2824860H1	SNP00024403	227	216	C	C	T	T57	n/a	n/a	n/a	n/a
79	7506079	4072670H1	SNP00024404	84	1391	G	G	C	noncoding	n/a	n/a	n/a	n/a
79	7506079	4255908H1	SNP00024403	16	215	C	C	T	H57	n/a	n/a	n/a	n/a
79	7506079	4365669H1	SNP00024404	181	1392	G	G	C	noncoding	n/a	n/a	n/a	n/a
79	7506079	4931664H1	SNP00053025	170	808	C	C	A	R254	0.96	n/d	n/d	n/d
79	7506079	5997473H1	SNP00053025	159	805	C	C	A	N253	0.96	n/d	n/d	n/d
81	7509263	032775H1	SNP00002923	217	704	T	C	T	noncoding	n/a	n/a	n/a	n/a
81	7509263	1833713H1	SNP00021247	183	1412	C	C	T	noncoding	n/a	n/a	n/a	n/a
81	7509263	2474113H1	SNP00002923	144	705	C	C	T	noncoding	n/a	n/a	n/a	n/a
81	7509263	2589335H1	SNP00021247	177	1409	C	C	T	noncoding	n/a	n/a	n/a	n/a
81	7509263	2682376H1	SNP00002923	201	703	C	C	T	noncoding	n/a	n/a	n/a	n/a
81	7509263	3234207H1	SNP00002923	192	701	C	C	T	noncoding	n/a	n/a	n/a	n/a
81	7509263	3818539H1	SNP00002923	41	702	T	C	T	noncoding	n/a	n/a	n/a	n/a
81	7509263	4227472H1	SNP00021247	36	1411	C	C	T	noncoding	n/a	n/a	n/a	n/a

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
81	7509263	4729911H1	SNP00021247	250	1410	C	C	T	noncoding	n/a	n/a	n/a	n/a
81	7509263	5317509H1	SNP00002923	190	678	T	C	T	noncoding	n/a	n/a	n/a	n/a
81	7509263	5872764H1	SNP00002923	125	697	T	C	T	noncoding	n/a	n/a	n/a	n/a
81	7509263	6740123H1	SNP00021247	433	1405	C	C	T	noncoding	n/a	n/a	n/a	n/a
82	7509360	099540H1	SNP00020576	40	1170	C	C	A	noncoding	n/d	n/d	n/d	n/d
82	7509360	1418649H1	SNP00099841	224	1120	T	T	C	noncoding	n/a	n/a	n/a	n/a
82	7509360	1418649H1	SNP00140985	26	922	A	A	G	noncoding	n/a	n/a	n/a	n/a
82	7509360	2562865H1	SNP00140984	129	548	A	A	G	M159	n/a	n/a	n/a	n/a
82	7509360	2880623H1	SNP00140984	198	546	A	A	G	E158	n/a	n/a	n/a	n/a
82	7509360	2958456H1	SNP00020576	46	1166	C	C	A	noncoding	n/d	n/d	n/d	n/d
82	7509360	2960578H1	SNP00020576	139	1167	C	C	A	noncoding	n/d	n/d	n/d	n/d
82	7509360	2960578H1	SNP00099841	89	1117	T	T	C	noncoding	n/a	n/a	n/a	n/a
82	7509360	3495068H1	SNP00020576	252	1169	C	C	A	noncoding	n/d	n/d	n/d	n/d
82	7509360	3495068H1	SNP00099841	202	1119	T	T	C	noncoding	n/a	n/a	n/a	n/a
82	7509360	3495068H1	SNP00140985	4	921	A	A	G	noncoding	n/a	n/a	n/a	n/a
82	7509360	3999001H1	SNP00020576	122	1171	C	C	A	noncoding	n/d	n/d	n/d	n/d
82	7509360	7017565H1	SNP00118148	49	1066	G	G	C	noncoding	n/d	n/d	n/a	n/d
82	7509360	7649114H2	SNP00099840	501	757	C	C	T	noncoding	n/a	n/a	n/a	n/a
83	7509394	1254595H1	SNP00037748	149	1190	T	C	T	noncoding	n/a	n/a	n/a	n/a
83	7509394	1379230H1	SNP00047004	80	1398	C	T	C	noncoding	n/a	n/a	n/a	n/a
83	7509394	1394251H1	SNP00037746	106	102	G	G	A	T29	n/d	n/d	n/d	n/d
83	7509394	1459712H1	SNP00076223	128	1760	C	C	T	noncoding	n/a	n/a	n/a	n/a
83	7509394	1542369H1	SNP00120103	167	2021	C	C	T	noncoding	n/d	n/a	n/a	n/a
83	7509394	1599310H1	SNP00009550	112	1606	G	A	G	noncoding	n/a	n/a	n/a	n/a
83	7509394	1625901H1	SNP00120103	121	2020	C	C	T	noncoding	n/d	n/a	n/a	n/a
83	7509394	1961909H1	SNP00009550	251	1604	G	A	G	noncoding	n/a	n/a	n/a	n/a
83	7509394	2193560H1	SNP00037749	227	2141	C	C	G	noncoding	n/a	n/a	n/a	n/a
83	7509394	2267960H1	SNP00076222	34	1307	C	C	T	noncoding	n/a	n/a	n/a	n/a

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
83	7509394	2293535H1	SNP00009551	208	2239	G	G	A	noncoding	n/a	n/a	n/a	n/a
83	7509394	2293535H1	SNP00037749	109	2140	C	C	G	noncoding	n/a	n/a	n/a	n/a
83	7509394	2865101H1	SNP00037746	104	98	G	G	A	R28	n/d	n/d	n/d	n/d
83	7509394	2870327H1	SNP00076223	123	1758	C	C	T	noncoding	n/a	n/a	n/a	n/a
83	7509394	2953469H1	SNP00037749	227	2131	C	C	G	noncoding	n/a	n/a	n/a	n/a
83	7509394	2953469H1	SNP00120103	108	2012	C	C	T	noncoding	n/d	n/a	n/a	n/a
83	7509394	3048001H1	SNP00120103	164	2018	C	C	T	noncoding	n/d	n/a	n/a	n/a
83	7509394	3114317H1	SNP00076222	145	1311	C	C	T	noncoding	n/a	n/a	n/a	n/a
83	7509394	3245235H1	SNP00037746	100	101	G	G	A	R29	n/d	n/d	n/d	n/d
83	7509394	3415619H1	SNP00009549	101	1344	C	C	T	noncoding	n/a	n/a	n/a	n/a
83	7509394	3537156H1	SNP00009549	11	1342	C	C	T	noncoding	n/a	n/a	n/a	n/a
83	7509394	3541107H1	SNP00037746	103	99	G	G	A	M28	n/d	n/d	n/d	n/d
83	7509394	3550907H1	SNP00076223	47	1759	C	C	T	noncoding	n/a	n/a	n/a	n/a
83	7509394	3833004H1	SNP00047004	143	1385	C	T	C	noncoding	n/a	n/a	n/a	n/a
83	7509394	3833004H1	SNP00076222	52	1294	C	C	T	noncoding	n/a	n/a	n/a	n/a
83	7509394	4098841H1	SNP00037746	93	100	G	G	A	A29	n/d	n/d	n/d	n/d
83	7509394	4177268H1	SNP00009549	106	1343	T	C	T	noncoding	n/a	n/a	n/a	n/a
83	7509394	4284807H1	SNP00120103	149	2016	C	C	T	noncoding	n/d	n/a	n/a	n/a
83	7509394	4302009H1	SNP00009550	99	1603	G	A	G	noncoding	n/a	n/a	n/a	n/a
83	7509394	4302009H1	SNP00076223	252	1757	C	C	T	noncoding	n/a	n/a	n/a	n/a
83	7509394	4591035H1	SNP00047004	28	1396	C	T	C	noncoding	n/a	n/a	n/a	n/a
83	7509394	4629540H1	SNP00120103	257	2019	C	C	T	noncoding	n/d	n/a	n/a	n/a
83	7509394	4689567H1	SNP00076222	172	1305	C	C	T	noncoding	n/a	n/a	n/a	n/a
83	7509394	4796307H1	SNP00047004	27	1395	C	T	C	noncoding	n/a	n/a	n/a	n/a
83	7509394	4993753H1	SNP00037746	89	87	G	G	A	S24	n/d	n/d	n/d	n/d
83	7509394	5019368H1	SNP00047004	102	1397	C	T	C	noncoding	n/a	n/a	n/a	n/a
83	7509394	5273937H1	SNP00047004	207	1400	C	T	C	noncoding	n/a	n/a	n/a	n/a
83	7509394	5628715H1	SNP00037746	77	86	G	G	A	C24	n/d	n/d	n/d	n/d

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
83	7509394	5786877H1	SNP00037749	240	2139	C	C	G	noncoding	n/a	n/a	n/a	n/a
83	7509394	5956252H1	SNP00037747	34	402	G	A	G	P129	n/a	n/a	n/a	n/a
83	7509394	6252652H1	SNP00076223	154	1755	C	C	T	noncoding	n/a	n/a	n/a	n/a
83	7509394	6451958H2	SNP00076221	162	593	C	C	T	noncoding	n/d	n/d	n/d	n/a
83	7509394	6538886H1	SNP00009551	258	2240	G	G	A	noncoding	n/a	n/a	n/a	n/a
83	7509394	680468511	SNP00037746	102	85	G	G	A	A24	n/d	n/d	n/d	n/d
83	7509394	691031711	SNP00076221	145	585	C	C	T	noncoding	n/d	n/d	n/d	n/a
83	7509394	7438966H1	SNP00037748	221	1186	C	C	T	noncoding	n/a	n/a	n/a	n/a
83	7509394	7644112H1	SNP00092752	290	902	C	C	G	noncoding	n/d	n/d	n/d	n/d
83	7509394	7690052H1	SNP00009549	156	1345	C	C	T	noncoding	n/a	n/a	n/a	n/a
83	7509394	776639H1	SNP00120103	222	2017	C	C	T	noncoding	n/d	n/a	n/a	n/a
85	7504551	075544H1	SNP00054749	29	962	C	C	G	noncoding	0.90	n/a	n/a	n/a
85	7504551	1476562H1	SNP00021336	105	431	G	G	C	noncoding	n/a	n/a	n/a	n/a
85	7504551	1954526H1	SNP00021336	81	393	G	G	C	noncoding	n/a	n/a	n/a	n/a
85	7504551	2095683H1	SNP00002980	56	73	A	A	C	T20	0.16	n/a	n/a	n/a
85	7504551	2823112H1	SNP00054749	215	960	C	C	G	noncoding	0.90	n/a	n/a	n/a
85	7504551	312505H1	SNP00054749	56	958	C	C	G	noncoding	0.90	n/a	n/a	n/a
85	7504551	312966H1	SNP00054749	57	959	C	C	G	noncoding	0.90	n/a	n/a	n/a
85	7504551	3490640H1	SNP00021336	29	429	G	G	C	noncoding	n/a	n/a	n/a	n/a
85	7504551	3505646H1	SNP00021336	71	428	G	G	C	noncoding	n/a	n/a	n/a	n/a
85	7504551	3714982H1	SNP00145418	156	939	T	G	T	noncoding	n/a	n/a	n/a	n/a
85	7504551	372809H1	SNP00054749	55	957	C	C	G	noncoding	0.90	n/a	n/a	n/a
85	7504551	3856868H1	SNP00054749	181	956	C	C	G	noncoding	0.90	n/a	n/a	n/a
85	7504551	3934724H1	SNP00021336	205	426	G	G	C	noncoding	n/a	n/a	n/a	n/a
85	7504551	4086175H1	SNP00054749	36	961	C	C	G	noncoding	0.90	n/a	n/a	n/a
85	7504551	4896753H1	SNP00021336	217	430	C	G	C	noncoding	n/a	n/a	n/a	n/a
85	7504551	5110335H1	SNP00054749	11	947	C	C	G	noncoding	0.90	n/a	n/a	n/a
85	7504551	5266431H1	SNP00054749	34	914	C	C	G	noncoding	0.90	n/a	n/a	n/a

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
85	7504551	5652679H1	SNP00145418	204	941	G	G	T	noncoding	n/a	n/a	n/a	n/a
85	7504551	6409965H1	SNP00021336	335	425	G	G	C	noncoding	n/a	n/a	n/a	n/a
85	7504551	6456689H1	SNP00021336	30	402	G	G	C	noncoding	n/a	n/a	n/a	n/a
85	7504551	6737774H1	SNP00054749	412	908	C	C	G	noncoding	0.90	n/a	n/a	n/a
86	7500652	1333863H1	SNP00024912	208	964	C	C	G	noncoding	n/a	n/a	n/a	n/a
86	7500652	1363519H1	SNP00024912	74	965	C	C	G	noncoding	n/a	n/a	n/a	n/a
86	7500652	1419657H1	SNP00018614	200	1459	C	C	T	noncoding	n/a	n/a	n/a	n/a
86	7500652	1451266H1	SNP00093232	173	688	C	C	T	noncoding	n/a	n/a	n/a	n/a
86	7500652	2208343H1	SNP00126814	164	1243	C	C	G	noncoding	n/a	n/a	n/a	n/a
86	7500652	2208763H1	SNP00126814	164	1244	C	C	G	noncoding	n/a	n/a	n/a	n/a
86	7500652	2211634H1	SNP00024912	14	963	C	C	G	noncoding	n/a	n/a	n/a	n/a
86	7500652	2579995H1	SNP00126814	282	1245	C	C	G	noncoding	n/a	n/a	n/a	n/a
86	7500652	2682565H1	SNP00024912	222	966	C	C	G	noncoding	n/a	n/a	n/a	n/a
86	7500652	3201709H1	SNP00024912	32	961	C	C	G	noncoding	n/a	n/a	n/a	n/a
86	7500652	4443586H1	SNP00093232	17	686	C	C	T	noncoding	n/a	n/a	n/a	n/a
86	7500652	5271544H1	SNP00126814	182	1246	C	C	G	noncoding	n/a	n/a	n/a	n/a
86	7500652	5795096H1	SNP00126814	141	1241	C	C	G	noncoding	n/a	n/a	n/a	n/a
86	7500652	5822746H1	SNP00018614	201	1457	C	C	T	noncoding	n/a	n/a	n/a	n/a
86	7500652	6213967H1	SNP00018614	40	1455	C	C	T	noncoding	n/a	n/a	n/a	n/a
86	7500652	6317066H1	SNP00018614	73	1456	C	C	T	noncoding	n/a	n/a	n/a	n/a
86	7500652	6495958H1	SNP00076191	162	78	A	G	A	noncoding	n/a	n/a	n/a	n/a
86	7500652	6498208H1	SNP00076191	162	76	G	G	A	noncoding	n/a	n/a	n/a	n/a
86	7500652	6700370H1	SNP00076191	31	112	G	G	A	noncoding	n/a	n/a	n/a	n/a
86	7500652	6821938J1	SNP00076191	161	70	G	G	A	noncoding	n/a	n/a	n/a	n/a
86	7500652	6824984H1	SNP00076191	168	103	G	G	A	noncoding	n/a	n/a	n/a	n/a
86	7500652	7195426H1	SNP00043172	206	852	A	A	C	noncoding	n/d	n/a	n/a	n/a
86	7500652	978784H1	SNP00024912	187	946	C	C	G	noncoding	n/a	n/a	n/a	n/a
87	7500900	1321148H1	SNP00058613	29	2538	T	T	C	noncoding	n/d	n/a	n/a	n/a

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
87	7500900	1321148H1	SNP00058614	77	2588	A	A	G	noncoding	n/d	n/d	n/d	n/d
87	7500900	1918601H1	SNP00116768	149	708	G	G	T	G168	n/d	n/d	n/d	n/d
87	7500900	2061322H1	SNP00116769	166	1436	T	T	C	noncoding	n/a	n/a	n/a	n/a
87	7500900	2299923H1	SNP00058614	51	2589	A	A	G	noncoding	n/d	n/d	n/d	n/d
87	7500900	2912334H1	SNP00058614	76	2585	G	A	G	noncoding	n/d	n/d	n/d	n/d
87	7500900	2945962H1	SNP00116768	179	704	G	G	T	V167	n/d	n/d	n/d	n/d
87	7500900	4151024H1	SNP00058613	39	2535	T	T	C	noncoding	n/d	n/a	n/a	n/a
87	7500900	4151024H1	SNP00058614	92	2590	A	A	G	noncoding	n/d	n/d	n/d	n/d
87	7500900	4210444H1	SNP00058613	4	2536	T	T	C	noncoding	n/d	n/a	n/a	n/a
87	7500900	426051H1	SNP00116768	152	707	G	G	T	E168	n/d	n/d	n/d	n/d
87	7500900	4653464H1	SNP00116769	138	1437	T	T	C	noncoding	n/a	n/a	n/a	n/a
87	7500900	4797905H1	SNP00116768	10	706	G	G	T	V167	n/d	n/d	n/d	n/d
87	7500900	5308481H1	SNP00116769	132	1435	T	T	C	noncoding	n/a	n/a	n/a	n/a
87	7500900	6368050H1	SNP00116769	265	1438	T	T	C	noncoding	n/a	n/a	n/a	n/a
87	7500900	64839H1	SNP00058614	58	2596	A	A	G	noncoding	n/d	n/d	n/d	n/d
88	7501398	1563261H1	SNP00059154	36	570	T	T	G	noncoding	n/a	n/a	n/a	n/a
88	7501398	3407970H1	SNP00059154	224	569	T	T	G	noncoding	n/a	n/a	n/a	n/a
88	7501398	640919H1	SNP00059153	67	460	C	C	T	A135	n/a	n/a	n/a	n/a
88	7501398	6568934H1	SNP00107591	387	1256	C	C	G	noncoding	n/a	n/a	n/a	n/a
89	7501417	6147633H1	SNP00049828	139	1343	G	G	A	noncoding	n/a	n/a	n/a	n/a
90	7501472	2675922H1	SNP00030909	26	290	G	C	G	R91	0.88	n/d	n/d	0.92
90	7501472	3532586H1	SNP00030909	62	289	C	C	G	P91	0.88	n/d	n/d	0.92
92	7501555	1281240H1	SNP00028378	29	583	G	G	C	G151	n/d	n/d	n/d	n/d
92	7501555	1431686H1	SNP00043008	126	980	T	C	T	noncoding	n/a	n/a	n/a	n/a
92	7501555	2395269H1	SNP00028378	137	584	G	G	C	S151	n/d	n/d	n/d	n/d
92	7501555	2659985H1	SNP00005601	84	820	A	A	G	T230	0.87	0.96	0.97	0.94
92	7501555	4120069H1	SNP00028378	61	582	G	G	C	L150	n/d	n/d	n/d	n/d
92	7501555	4570910H1	SNP00028378	241	578	G	G	C	G149	n/d	n/d	n/d	n/d

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
92	7501555	4988663H1	SNP00043008	138	983	C	C	T	noncoding	n/a	n/a	n/a	n/a
92	7501555	5320593H1	SNP00028378	154	576	G	G	C	A148	n/d	n/d	n/d	n/d
92	7501555	5872846H1	SNP00028378	24	406	G	G	C	G92	n/d	n/d	n/d	n/d
93	7501561	6369074H1	SNP00058179	385	382	C	C	G	noncoding	n/d	n/a	n/a	n/a
94	7506108	016157H1	SNP00130459	79	1360	C	C	T	noncoding	n/a	n/a	n/a	n/a
94	7506108	1403510H1	SNP00139261	23	455	G	G	C	L35	n/a	n/a	n/a	n/a
94	7506108	1597068H1	SNP00124524	124	404	T	T	G	F18	n/a	n/a	n/a	n/a
94	7506108	2131028H1	SNP00130459	83	1356	C	C	T	noncoding	n/a	n/a	n/a	n/a
94	7506108	2331002H1	SNP00130459	186	1353	C	C	T	noncoding	n/a	n/a	n/a	n/a
94	7506108	2512454H1	SNP00130459	95	1359	C	C	T	noncoding	n/a	n/a	n/a	n/a
94	7506108	2635234H1	SNP00124524	154	405	T	T	G	W19	n/a	n/a	n/a	n/a
94	7506108	3558448H1	SNP00124524	173	403	T	T	G	F18	n/a	n/a	n/a	n/a
94	7506108	3869239H1	SNP00130459	99	1358	C	C	T	noncoding	n/a	n/a	n/a	n/a
94	7506108	4085445H1	SNP00130459	54	1278	C	C	T	noncoding	n/a	n/a	n/a	n/a
94	7506108	4638567H1	SNP00139261	40	453	C	G	C	L35	n/a	n/a	n/a	n/a
94	7506108	5855320H1	SNP00130459	236	1355	C	C	T	noncoding	n/a	n/a	n/a	n/a
94	7506108	5887432H1	SNP00130459	44	1357	C	C	T	noncoding	n/a	n/a	n/a	n/a
94	7506108	6105187H1	SNP00130459	94	1352	C	C	T	noncoding	n/a	n/a	n/a	n/a
95	7506123	1229348H1	SNP00001180	56	56	C	C	G	noncoding	0.26	0.14	0.49	0.49
95	7506123	1230239H1	SNP00016950	211	1643	A	A	G	noncoding	n/a	n/a	n/a	n/a
95	7506123	1232226H1	SNP00055317	77	1053	C	T	C	noncoding	0.06	0.36	n/d	0.22
95	7506123	1232279H1	SNP00055317	78	1054	C	T	C	noncoding	0.06	0.36	n/d	0.22
95	7506123	2757058H1	SNP00016950	71	1544	T	T	C	noncoding	n/a	n/a	n/a	n/a
95	7506123	2967768H1	SNP00001180	44	53	C	C	G	noncoding	0.26	0.14	0.49	0.49
95	7506123	2984501H1	SNP00001180	48	41	G	C	G	noncoding	0.26	0.14	0.49	0.49
95	7506123	3297637H1	SNP00001180	55	54	C	C	G	noncoding	0.26	0.14	0.49	0.49
95	7506123	3462862H1	SNP00001180	61	55	G	C	G	noncoding	0.26	0.14	0.49	0.49
95	7506123	3591274H1	SNP00001180	53	48	G	C	G	noncoding	0.26	0.14	0.49	0.49

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
95	7506123	3731344H1	SNP00016950	93	1644	A	A	G	noncoding	n/a	n/a	n/a	n/a
95	7506123	3973791H1	SNP00001180	56	52	C	C	G	noncoding	0.26	0.14	0.49	0.49
95	7506123	4512701H1	SNP00016950	174	1639	A	A	G	noncoding	n/a	n/a	n/a	n/a
95	7506123	5663612H1	SNP00055317	55	1055	C	T	C	noncoding	0.06	0.36	n/d	0.22
95	7506123	5860537H1	SNP00001180	41	49	C	C	G	noncoding	0.26	0.14	0.49	0.49
95	7506123	6368475H1	SNP00016950	70	1641	A	A	G	noncoding	n/a	n/a	n/a	n/a
96	7506248	061878H1	SNP00104714	74	3165	A	A	G	K1008	n/a	n/a	n/a	n/a
96	7506248	2435340H1	SNP00069056	164	3280	C	C	G	A1046	n/a	n/a	n/a	n/a
96	7506248	2508587H1	SNP00124435	65	719	G	G	A	E193	n/d	n/d	n/d	n/d
96	7506248	3465435H1	SNP00025569	105	1450	C	C	T	N436	n/a	n/a	n/a	n/a
96	7506248	4313356H1	SNP00104714	197	3163	A	A	G	K1007	n/a	n/a	n/a	n/a
96	7506248	4553855H1	SNP00025569	199	1453	C	C	T	F437	n/a	n/a	n/a	n/a
96	7506248	4990154H1	SNP00025569	139	1449	T	C	T	I436	n/a	n/a	n/a	n/a
96	7506248	5264750H2	SNP00104714	120	3164	A	A	G	K1008	n/a	n/a	n/a	n/a
96	7506248	5919737H1	SNP00124436	178	1044	C	C	T	P301	n/d	n/a	n/a	n/a
96	7506248	5928396H1	SNP00069056	261	3247	C	C	G	R1035	n/a	n/a	n/a	n/a
96	7506248	5928396H1	SNP00104714	146	3132	A	A	G	E997	n/a	n/a	n/a	n/a
96	7506248	6170507H1	SNP00104714	198	3162	A	A	G	K1007	n/a	n/a	n/a	n/a
96	7506248	6249921H1	SNP00069056	233	3279	C	C	G	A1046	n/a	n/a	n/a	n/a
96	7506248	6378645H1	SNP00104714	176	3160	A	A	G	S1006	n/a	n/a	n/a	n/a
96	7506248	6465731H1	SNP00025569	326	1424	C	C	T	P428	n/a	n/a	n/a	n/a
96	7506248	6515577H1	SNP00124436	373	1047	C	C	T	A302	n/d	n/a	n/a	n/a
96	7506248	6779923J1	SNP00124436	94	1040	C	C	T	Q300	n/d	n/a	n/a	n/a
96	7506248	6908561J1	SNP00124435	114	711	G	G	A	S190	n/d	n/d	n/d	n/d
96	7506248	6908561J1	SNP00124436	442	1039	C	C	T	N299	n/d	n/a	n/a	n/a
96	7506248	7125058H1	SNP00069056	335	3232	C	C	G	C1030	n/a	n/a	n/a	n/a
96	7506248	7125058H1	SNP00104714	220	3117	A	A	G	H992	n/a	n/a	n/a	n/a
96	7506248	7125158H1	SNP00069056	353	3277	C	C	G	G1045	n/a	n/a	n/a	n/a

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
96	7506248	8533289H1	SNP00069056	246	3298	C	C	G	V1052	n/a	n/a	n/a	n/a
96	7506248	8533289H1	SNP00104714	131	3183	A	A	G	N1014	n/a	n/a	n/a	n/a
96	7506248	8608853J1	SNP00069056	495	3282	C	C	G	P1047	n/a	n/a	n/a	n/a
96	7506248	8608853J1	SNP00104714	380	3167	A	A	G	T1009	n/a	n/a	n/a	n/a
97	7506347	2540112H1	SNP00032080	78	1268	G	G	A	A386	n/a	n/a	n/a	n/a
97	7506347	2667339H1	SNP00092775	144	1741	C	C	T	noncoding	n/a	n/a	n/a	n/a
97	7506347	7610890J1	SNP00032079	312	1123	A	G	A	V337	n/a	n/a	n/a	n/a
97	7506347	7612418H1	SNP00032082	93	1794	G	G	A	noncoding	0.94	0.99	0.94	0.94
98	7509172	086872H1	SNP00073588	43	1489	G	A	G	noncoding	n/a	n/a	n/a	n/a
98	7509172	4419451H1	SNP00027681	137	683	C	C	T	noncoding	n/a	n/a	n/a	n/a
98	7509172	5396974H1	SNP00043998	221	1790	C	C	T	noncoding	n/a	n/a	n/a	n/a
98	7509172	6895504H1	SNP00110389	499	670	G	G	A	noncoding	0.97	n/d	n/d	n/d
98	7509172	6899923H1	SNP00110389	476	663	G	G	A	noncoding	0.97	n/d	n/d	n/d
99	7510421	075207H1	SNP00111604	114	1158	C	C	T	noncoding	n/a	n/a	n/a	n/a
99	7510421	1442920H1	SNP00111604	118	1162	C	C	T	noncoding	n/a	n/a	n/a	n/a
99	7510421	1547515H1	SNP00111604	56	1161	C	C	T	noncoding	n/a	n/a	n/a	n/a
99	7510421	1707635H1	SNP00119588	100	605	T	T	G	noncoding	n/d	n/d	n/d	n/d
99	7510421	1830039H1	SNP00021209	25	856	C	C	T	noncoding	n/a	n/a	n/a	n/a
99	7510421	3483358H1	SNP00021209	108	852	C	C	T	noncoding	n/a	n/a	n/a	n/a
99	7510421	3495367H1	SNP00111604	15	1160	C	C	T	noncoding	n/a	n/a	n/a	n/a
99	7510421	3602878H1	SNP00021209	8	855	C	C	T	noncoding	n/a	n/a	n/a	n/a
99	7510421	4509882H1	SNP00092952	61	1584	G	A	G	noncoding	n/a	n/a	n/a	n/a
99	7510421	4663322H1	SNP00111604	114	1159	C	C	T	noncoding	n/a	n/a	n/a	n/a
99	7510421	4884128H1	SNP00119588	113	602	G	T	G	noncoding	n/d	n/d	n/d	n/d
99	7510421	5802838H1	SNP00111604	90	1157	C	C	T	noncoding	n/a	n/a	n/a	n/a
99	7510421	6866675H1	SNP00092952	126	1586	G	A	G	noncoding	n/a	n/a	n/a	n/a
99	7510421	967018H1	SNP00111604	234	1163	C	C	T	noncoding	n/a	n/a	n/a	n/a
100	7504625	1304767H1	SNP00001992	113	206	A	A	G	N31	n/a	n/a	n/a	n/a

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
100	7504625	1304767H1	SNP00140101	101	194	C	C	T	R27	n/a	n/a	n/a	n/a
100	7504625	3742874H1	SNP00140101	89	143	C	C	T	Q10	n/a	n/a	n/a	n/a
100	7504625	4216671H1	SNP00001992	80	205	G	A	G	V30	n/a	n/a	n/a	n/a
100	7504625	4216671H1	SNP00140101	68	193	C	C	T	V26	n/a	n/a	n/a	n/a
100	7504625	5772720H1	SNP00140100	26	74	G	A	G	noncoding	n/a	n/a	n/a	n/a
101	7504776	099563H1	SNP00029752	111	791	A	A	G	T246	n/a	n/a	n/a	n/a
101	7504776	099563H1	SNP00101027	55	735	A	A	G	Y227	n/a	n/a	n/a	n/a
101	7504776	1379082H1	SNP00029752	153	802	A	A	G	Q249	n/a	n/a	n/a	n/a
101	7504776	1379082H1	SNP00101027	97	746	A	A	G	K231	n/a	n/a	n/a	n/a
101	7504776	1417737H1	SNP00029752	109	803	A	A	G	N250	n/a	n/a	n/a	n/a
101	7504776	1417737H1	SNP00101027	53	747	A	A	G	E231	n/a	n/a	n/a	n/a
101	7504776	1417910H1	SNP00101026	15	588	A	A	G	E178	n/a	n/a	n/a	n/a
101	7504776	1418992H1	SNP00101025	110	447	A	A	G	D131	n/d	n/d	n/d	n/d
101	7504776	221245H1	SNP00101027	212	748	A	A	G	E231	n/a	n/a	n/a	n/a
101	7504776	2365849H1	SNP00029752	218	805	A	A	G	K250	n/a	n/a	n/a	n/a
101	7504776	2367556H1	SNP00029751	7	22	T	T	C	noncoding	n/a	n/a	n/a	n/a
101	7504776	2561872H1	SNP00006435	20	194	T	C	T	C47	0.47	n/a	n/a	n/a
101	7504776	2561872H1	SNP00006436	77	251	T	G	T	F66	n/a	n/a	n/a	n/a
101	7504776	2664571H1	SNP00029752	163	801	A	A	G	H249	n/a	n/a	n/a	n/a
101	7504776	2664571H1	SNP00101027	107	745	A	A	G	G230	n/a	n/a	n/a	n/a
101	7504776	2961063H1	SNP00029751	6	21	T	T	C	noncoding	n/a	n/a	n/a	n/a
101	7504776	3501625H1	SNP00101025	287	444	A	A	G	Q130	n/d	n/d	n/d	n/d
101	7504776	3503710H1	SNP00101025	38	445	A	A	G	Q130	n/d	n/d	n/d	n/d
101	7504776	3503710H1	SNP00101026	179	586	A	A	G	stop177	n/a	n/a	n/a	n/a
101	7504776	3504402H1	SNP00101025	37	446	A	A	G	N131	n/d	n/d	n/d	n/d
101	7504776	3504402H1	SNP00101026	178	587	A	A	G	K178	n/a	n/a	n/a	n/a
101	7504776	4073560H1	SNP00101027	267	744	A	A	G	D230	n/a	n/a	n/a	n/a
101	7504776	4159304H1	SNP00101026	118	579	A	A	G	K175	n/a	n/a	n/a	n/a

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
101	7504776	6812405J1	SNP00101022	87	106	G	G	A	W17	n/d	n/d	n/d	n/d
101	7504776	6812405J1	SNP00132939	42	62	G	G	A	E3	n/a	n/a	n/a	n/a
101	7504776	7650236H2	SNP00006435	156	193	C	C	T	H46	0.47	n/a	n/a	n/a
101	7504776	7650236H2	SNP00006436	99	250	G	G	T	S65	n/a	n/a	n/a	n/a
101	7504776	835709H1	SNP00006436	62	275	T	G	T	L74	n/a	n/a	n/a	n/a
101	7504776	941474H1	SNP00101026	285	596	A	A	G	N181	n/a	n/a	n/a	n/a
102	7504927	1801846H1	SNP00011376	119	403	C	C	T	P108	n/a	n/a	n/a	n/a
102	7504927	2564524H1	SNP00152645	164	347	C	C	T	T89	n/a	n/a	n/a	n/a
102	7504927	2568409H1	SNP00152645	32	344	C	C	T	A88	n/a	n/a	n/a	n/a
102	7504927	3809704H1	SNP00011376	161	401	C	C	T	P107	n/a	n/a	n/a	n/a
102	7504927	3811993H1	SNP00011376	155	395	C	C	T	A105	n/a	n/a	n/a	n/a
102	7504927	4137230H1	SNP00011376	274	402	C	C	T	P107	n/a	n/a	n/a	n/a
102	7504927	4300762H1	SNP00011376	151	400	C	C	T	P107	n/a	n/a	n/a	n/a
102	7504927	5575359H1	SNP00152645	134	288	C	C	T	D69	n/a	n/a	n/a	n/a
102	7504927	5994671H1	SNP00011376	100	398	C	C	T	T106	n/a	n/a	n/a	n/a
102	7504927	6411426H1	SNP00011376	472	399	C	C	T	T106	n/a	n/a	n/a	n/a
102	7504927	8630325J1	SNP00011376	505	415	C	C	T	L112	n/a	n/a	n/a	n/a
103	7505010	001021H1	SNP00030777	223	1401	G	G	T	noncoding	n/d	n/a	n/a	n/a
103	7505010	001086H1	SNP00030777	190	1427	G	G	T	noncoding	n/d	n/a	n/a	n/a
103	7505010	030086H1	SNP00016835	42	975	T	T	C	L286	n/a	n/a	n/a	n/a
103	7505010	070827H1	SNP00030777	76	1428	G	G	T	noncoding	n/d	n/a	n/a	n/a
103	7505010	089673H1	SNP00098113	40	246	A	A	G	K43	n/a	n/a	n/a	n/a
103	7505010	1243032H1	SNP00060996	29	743	C	C	T	Q209	n/a	n/a	n/a	n/a
103	7505010	1274846H1	SNP00098113	69	247	A	A	G	K43	n/a	n/a	n/a	n/a
103	7505010	166695H1	SNP00060996	187	740	C	C	T	P208	n/a	n/a	n/a	n/a
103	7505010	2209369H1	SNP00060996	63	742	C	C	T	P208	n/a	n/a	n/a	n/a
103	7505010	2258875H1	SNP00016835	64	973	T	T	C	L285	n/a	n/a	n/a	n/a
103	7505010	2416727H1	SNP00098113	132	245	A	A	G	K43	n/a	n/a	n/a	n/a

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
103	7505010	2513994H2	SNP00098113	207	248	A	A	G	T44	n/a	n/a	n/a	n/a
103	7505010	3031953H1	SNP00030777	164	1425	T	G	T	noncoding	n/d	n/a	n/a	n/a
103	7505010	3078471H1	SNP00098113	76	238	A	A	G	stop40	n/a	n/a	n/a	n/a
103	7505010	3081869H1	SNP00098113	87	244	A	A	G	G42	n/a	n/a	n/a	n/a
103	7505010	3315260H1	SNP00098113	176	241	A	A	G	E41	n/a	n/a	n/a	n/a
103	7505010	3465309H1	SNP00060996	203	741	C	C	T	P208	n/a	n/a	n/a	n/a
103	7505010	3590354H1	SNP00098113	34	234	A	A	G	stop39	n/a	n/a	n/a	n/a
103	7505010	3591719H1	SNP00030777	10	1424	G	G	T	noncoding	n/d	n/a	n/a	n/a
103	7505010	4284425H1	SNP00098113	86	243	A	A	G	D42	n/a	n/a	n/a	n/a
103	7505010	4957085H1	SNP00030777	126	1423	G	G	T	noncoding	n/d	n/a	n/a	n/a
103	7505010	6191693H1	SNP00098113	61	232	G	A	G	K38	n/a	n/a	n/a	n/a
103	7505010	6526052H1	SNP00060995	434	567	G	A	G	R150	n/a	n/a	n/a	n/a
103	7505010	6526052H1	SNP00098113	138	271	A	A	G	L51	n/a	n/a	n/a	n/a
103	7505010	6526152H1	SNP00060995	362	492	G	A	G	G125	n/a	n/a	n/a	n/a
103	7505010	6526152H1	SNP00098113	66	196	A	A	G	I26	n/a	n/a	n/a	n/a
103	7505010	6526280H1	SNP00060995	413	542	G	A	G	E142	n/a	n/a	n/a	n/a
103	7505010	6729710H1	SNP00060996	546	737	T	C	T	L207	n/a	n/a	n/a	n/a
103	7505010	6817646J1	SNP00030777	431	1426	G	G	T	noncoding	n/d	n/a	n/a	n/a
103	7505010	7648442H1	SNP00098114	474	1103	C	C	A	H329	n/a	n/a	n/a	n/a
105	7510061	1560250H1	SNP00107470	86	2109	A	A	G	S624	0.96	n/a	n/a	n/a
105	7510061	1784861H1	SNP00009615	192	3153	T	T	G	noncoding	n/a	n/a	n/a	n/a
105	7510061	2196685H1	SNP00037844	78	1879	G	G	T	R547	n/d	n/a	n/a	n/a
105	7510061	2350164H1	SNP00009614	99	3206	C	C	T	noncoding	0.99	0.97	0.97	n/a
105	7510061	2439943H1	SNP00107471	186	2259	G	G	A	E674	n/d	n/d	n/d	n/d
105	7510061	2554655H1	SNP00052935	166	856	C	C	G	A206	n/d	n/d	n/d	n/d
105	7510061	3112878H1	SNP00037844	31	1877	G	G	T	V546	n/d	n/a	n/a	n/a
105	7510061	3112878H1	SNP00107470	260	2107	A	A	G	Q623	0.96	n/a	n/a	n/a
105	7510061	3369632H1	SNP00107470	61	2108	A	A	G	L623	0.96	n/a	n/a	n/a

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
105	7510061	3369632H1	SNP00107471	211	2258	G	G	A	L673	n/d	n/d	n/d	n/d
105	7510061	3508437H1	SNP00009614	208	3203	C	C	T	noncoding	0.99	0.97	0.97	n/a
105	7510061	3508437H1	SNP00009615	155	3150	T	T	G	noncoding	n/a	n/a	n/a	n/a
105	7510061	3523022H1	SNP00009614	252	3205	C	C	T	noncoding	0.99	0.97	0.97	n/a
105	7510061	3523022H1	SNP00009615	199	3152	T	T	G	noncoding	n/a	n/a	n/a	n/a
105	7510061	3875558H1	SNP00009614	181	3204	C	C	T	noncoding	0.99	0.97	0.97	n/a
105	7510061	3875558H1	SNP00009615	128	3151	T	T	G	noncoding	n/a	n/a	n/a	n/a
105	7510061	6341477H1	SNP00120597	416	3342	A	G	A	noncoding	n/a	n/a	n/a	n/a
106	7510091	010579H1	SNP00020548	63	1939	T	T	C	noncoding	n/a	n/a	n/a	n/a
106	7510091	083781H1	SNP00020547	143	1288	T	T	C	noncoding	0.96	n/d	n/d	0.98
106	7510091	1510017H1	SNP00020548	66	1938	T	T	C	noncoding	n/a	n/a	n/a	n/a
106	7510091	3091941H1	SNP00020548	73	1931	T	T	C	noncoding	n/a	n/a	n/a	n/a
106	7510091	3520176H1	SNP00020547	169	1277	T	T	C	noncoding	0.96	n/d	n/d	0.98
106	7510091	3880313H1	SNP00020548	46	1936	T	T	C	noncoding	n/a	n/a	n/a	n/a
106	7510091	4726056H1	SNP00020547	159	1284	T	T	C	noncoding	0.96	n/d	n/d	0.98
106	7510091	866517H1	SNP00020548	117	1887	C	T	C	noncoding	n/a	n/a	n/a	n/a
107	7510109	3700051H1	SNP00118844	140	268	A	A	G	K22	n/d	n/a	n/a	n/a
108	7510121	007977H1	SNP00006647	139	1991	C	T	C	noncoding	n/a	n/a	n/a	n/a
108	7510121	1447072H1	SNP00135814	158	1853	A	A	G	noncoding	n/a	n/a	n/a	n/a
108	7510121	3351653H1	SNP00135814	147	1852	A	A	G	noncoding	n/a	n/a	n/a	n/a
108	7510121	3441378H1	SNP00006647	81	1992	T	T	C	noncoding	n/a	n/a	n/a	n/a
108	7510121	3839661H1	SNP00006647	80	1989	T	T	C	noncoding	n/a	n/a	n/a	n/a
108	7510121	3969395H1	SNP00006647	217	1990	T	T	C	noncoding	n/a	n/a	n/a	n/a
108	7510121	3969395H1	SNP00135814	79	1851	A	A	G	noncoding	n/a	n/a	n/a	n/a
108	7510121	3970869H1	SNP00135814	76	1849	A	A	G	noncoding	n/a	n/a	n/a	n/a
108	7510121	4404232H1	SNP00135814	150	1854	A	A	G	noncoding	n/a	n/a	n/a	n/a
109	7510797	2717927H1	SNP00067918	195	341	T	C	T	G109	n/a	n/a	n/a	n/a
109	7510797	3441302H1	SNP00099646	122	2544	T	T	C	noncoding	n/a	n/a	n/a	n/a

Table 8

SEQ ID NO:	PID	EST ID	SNP ID	EST SNP	CB1 SNP	EST Allele	Allele 1	Allele 2	Amino Acid	Caucasian Allele 1 frequency	African Allele 1 frequency	Asian Allele 1 frequency	Hispanic Allele 1 frequency
109	7510797	6755547J1	SNP00045198	96	2229	C	C	T	noncoding	0.90	n/a	n/a	n/a
109	7510797	6765874J1	SNP00099643	398	4	A	G	A	noncoding	n/a	n/a	n/a	n/a
109	7510797	7235093H1	SNP00099644	371	2223	G	G	A	noncoding	n/a	n/a	n/a	n/a
109	7510797	8530932H1	SNP00092880	368	2615	A	A	G	noncoding	n/a	n/a	n/a	n/a
110	7504944	1458467H1	SNP00033323	182	2005	A	A	G	noncoding	n/a	n/a	n/a	n/a
110	7504944	1478020H1	SNP00057452	115	1522	C	T	C	L498	n/a	n/a	n/a	n/a
110	7504944	2101051H1	SNP00107397	169	845	C	C	T	T272	n/d	n/d	n/d	n/d
110	7504944	2115787H1	SNP00033323	39	2004	A	A	G	noncoding	n/a	n/a	n/a	n/a
110	7504944	2186256H1	SNP00107397	210	842	C	C	T	A271	n/d	n/d	n/d	n/d
110	7504944	2218706H1	SNP00033323	82	2002	A	A	G	noncoding	n/a	n/a	n/a	n/a
110	7504944	2599123H1	SNP00145840	132	1402	A	A	G	N458	n/a	n/a	n/a	n/a
110	7504944	2599123H1	SNP00145841	139	1409	T	T	C	F460	n/a	n/a	n/a	n/a
110	7504944	2718732H1	SNP00007660	230	1823	T	T	C	L598	n/a	n/a	n/a	n/a
110	7504944	2923394H1	SNP00007659	27	1617	C	C	T	Y529	n/a	n/a	n/a	n/a
110	7504944	4349655H1	SNP00007660	180	1820	T	T	C	I597	n/a	n/a	n/a	n/a
110	7504944	4368856H1	SNP00007659	212	1614	C	C	T	N528	n/a	n/a	n/a	n/a
110	7504944	4606685H1	SNP00107397	167	837	C	C	T	V269	n/d	n/d	n/d	n/d
110	7504944	4864415H1	SNP00107397	170	841	C	C	T	P271	n/d	n/d	n/d	n/d
110	7504944	5185583H1	SNP00033323	177	1951	A	A	G	noncoding	n/a	n/a	n/a	n/a
110	7504944	5956324H1	SNP00007659	305	1612	C	C	T	Q528	n/a	n/a	n/a	n/a
110	7504944	6410128H1	SNP00007660	137	1818	T	T	C	N596	n/a	n/a	n/a	n/a
110	7504944	6453066H1	SNP00126677	37	1458	C	T	C	N476	n/a	n/a	n/a	n/a
110	7504944	8095428H1	SNP00047425	62	926	A	C	A	Y299	n/a	n/a	n/a	n/a
110	7504944	929739H1	SNP00057452	172	1520	T	T	C	V497	n/a	n/a	n/a	n/a